

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2004, 05:04:20 ; Search time 5952 Seconds

(without alignment)
11649.863 Million cell updates/sec

Title: US-10-074-744-4
Perfect score: 2322
Sequence: 1 aagccttctcaattcaacct.....gtcgtagctggtgacacaca 2322

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estbm:*
3: em_estcn:*
4: em_estcm:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_estcrl:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estcm:*
16: em_estcm:*
17: em_estcm:*
18: em_estcm:*
19: em_estcm:*
20: em_estcm:*
21: em_estcm:*
22: em_estcm:*
23: em_estcm:*
24: em_estcm:*
25: em_estcm:*
26: em_estcm:*
27: em_estcm:*
28: gb_estc1:*
29: gb_estc2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635.2	27.4	722	10	AM983873 RCI-HN000
2	628.6	27.1	649	10	AM983877 RCI-HN000
3	602.8	26.0	663	10	AM983875 RCI-HN000
4	497.6	21.4	547	10	AM984211 PM2-HN000

5	400.8	17.3	446	10	EG010614
6	314	13.5	336	14	R04703
7	309.6	13.3	365	10	BF737769
8	238.6	10.3	262	10	AM984213
9	205.8	8.9	254	10	BF773183
10	202.8	8.7	206	10	BF348952
11	66.2	2.9	1101	29	CNS0039G
12	65.8	2.8	1101	29	CNS0039G
13	65.4	2.8	556	10	BF755979
14	61	2.6	994	13	EX414650
15	59.8	2.6	961	29	CNS008HT
16	56.6	2.4	1201	13	EX355654
17	55.8	2.4	1146	29	CNS021G2
18	54	2.3	1201	13	EX461310
19	53.8	2.3	1091	13	EX424950
20	53.6	2.3	1124	13	EX436282
21	52.8	2.3	1101	29	CNS0106X
22	52.4	2.3	1204	29	CNS0106X
23	51.8	2.2	637	29	CNS036CC
24	51.2	2.2	897	29	CNS015FK
25	51	2.2	1201	9	AL565455
26	50.8	2.2	598	12	BJ435904
27	50.8	2.2	1044	13	EX415231
28	50.6	2.2	1101	29	CNS0176D
29	50.6	2.2	1101	29	CNS0182P
30	50.6	2.2	1173	13	EX443524
31	50.4	2.2	928	29	CNS000DY
32	50.2	2.2	1199	13	EX375702
33	50	2.2	1056	13	EX415058
34	49.8	2.1	1201	13	EX361152
35	49.4	2.1	697	28	BZ012209
36	49.4	2.1	1201	13	EX396546
37	49.2	2.1	987	29	CNS014PQ
38	49.2	2.1	1099	13	EX456575
39	49	2.1	1200	13	EX415878
40	48.8	2.1	1101	29	CNS0083U
41	48.6	2.1	1092	29	CNS020K7
42	48.6	2.1	1201	9	AL559324
43	48.4	2.1	1101	29	CNS0006J
44	48.2	2.1	1201	13	EX363491
45	48.2	2.1	1201	13	EX386038

ALIGNMENTS

RESULT 1
LOCUS AM983873 722 bp mRNA linear EST 02-JUN-2000
DEFINITION RCI-HN0003-220300-021-C03 HN0003 Homo sapiens CDNA, mRNA sequence.
ACCESSION AM983873
VERSION AM983873.1 GI:8175467
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 722)
AUTHORS Dias Neto E., Garcia Correa, R., Verjovski-Almeida, S., Britons, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
Sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=kt2=RC1-HN0003-220
 300-021-c03&ct3=2000-03-22&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 40
 High quality sequence stop: 685.
 Location/Qualifiers

FEATURES

source

1. 722
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HN0003"
 /note="Organ: head normal; Vector: puc18; Site 1: SmaI;
 Site 2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 27.4%; Score 635.2; DB 10; Length 722;

Best Local Similarity 98.6%; Pred. No. 2.3e-137; Indels 1; Gaps 1;
 Matches 651; Conservative 0; Mismatches 8;

```

6  TTCTTAATTAACCTTGTGAGGTTACCACTAAGTTAGGCTCAAGAGGGTGTG 65
28  TTCTTAATTAACCTTGTGAGGTTACCACTAAGTTAGGCTCAAGAGGGTGTG 86
66  TCTGTGTGAGGTTAACTGACCTGCGAAGTTAATTTCTAATTTGTTCTTCT 125
87  TCTGTGTGAGGTTAACTGACCTGCGAAGTTAATTTCTAATTTGTTCTTCT 146
126  GCAAAAAGGGGAGAGTGAATGAATTAATTTCTAATTTGTTCTTCTTCT 185
147  GCAAAAAGGGGAGAGTGAATGAATTAATTTCTAATTTGTTCTTCTTCT 206
186  CCGAGCAATTAATTAAGCAATTCGCTATAGTTCGCTGGAAGAGTAGTTTTCATG 245
207  CCGAGCAATTAATTAAGCAATTCGCTATAGTTCGCTGGAAGAGTAGTTTTCATG 266
246  TACTTAACCTTCACTCTGTTCAATTCATGCTTTTAAACGGTTGACCTTCTAATCC 305
267  TACTTAACCTTCACTCTGTTCAATTCATGCTTTTAAACGGTTGACCTTCTAATCC 326
306  TATCTGACCTTAATTAATTTTAAAGTGTTCATTAAGAAAGCTCTGATCAACGAGCTG 365
327  TATCTGACCTTAATTAATTTTAAAGTGTTCATTAAGAAAGCTCTGATCAACGAGCTG 386
366  CGATAATTAAGTGTGTATCCAGATTTGTCATCTCACTCACTCACTCACTCACTCACT 425
387  CGATAATTAAGTGTGTATCCAGATTTGTCATCTCACTCACTCACTCACTCACTCACT 446
426  AACACCTTAAGTGTGTATCCAGATTTGTCATCTCACTCACTCACTCACTCACTCACT 485
447  AACACCTTAAGTGTGTATCCAGATTTGTCATCTCACTCACTCACTCACTCACTCACT 506
486  TGACCGTAGAGCTTCCACATGAGATTTTGAACCTCTGACGATCTACTGAGGGAATG 545
507  TGACCGTAGAGCTTCCACATGAGATTTTGAACCTCTGACGATCTACTGAGGGAATG 566
546  AGTTGCAATTAATTTGTAACACCATGCGTGCATGAGTAGTGGCTTAATGTTGTAATA 605
567  AGTTGCAATTAATTTGTAACACCATGCGTGCATGAGTAGTGGCTTAATGTTGTAATA 626
606  AGCAAGAGCAAGAGTGTGATGAGTGAACCTCTGTTCAATGATTAATTAATGACTT 665

```

Db 627 AGCAGAGCAAGAGTGTGATGAGTGAACCTCTGTTCAATGATTAATTAATGACTT 686

RESULT 2
 LOCUS AM983877 649 bp mRNA linear EST 02-JUN-2000
 DEFINITION RCL-HN0003-220300-021-f03 HN0003 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AM983877
 VERSION AM983877.1 GI:8175471
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 649)
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zagro, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

PUBMED

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=kt2=RC1-HN0003-220
 300-021-f03&ct3=2000-03-22&ct4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 3
 High quality sequence stop: 646.
 Location/Qualifiers

FEATURES

source

1. 649
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="HN0003"
 /note="Organ: head normal; Vector: puc18; Site 1: SmaI;
 Site 2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

ORIGIN

Query Match 27.1%; Score 628.6; DB 10; Length 649;

Best Local Similarity 99.2%; Pred. No. 7.9e-136; Indels 1; Gaps 1;
 Matches 642; Conservative 0; Mismatches 4;

```

2  AGCTTTCTAATTAACCTTGTGAGGTTACCACTAAGTTAGGCTCAAGAGGG 61
4  AGCTTTCTAATTAACCTTGTGAGGTTACCACTAAGTTAGGCTCAAGAGGG 62
62  TGTGTCTGTGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 121
63  TGTGTCTGTGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 122
122  TTTCGCAAAAAGTGGGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 181
123  TTTCGCAAAAAGTGGGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 182

```

Location/Qualifiers
1. .663

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

QY 670 CCAAGACACTTACTAATCTGTGATAGTAATAAACAATTGCATGTCAGAGCTCATT 729
 |||||
 Db 63 CCAAGACACTTACTAATCTGTGATAGTAATAAACAATTGCATGTCAGAGCTCATT 122
 |||||
 QY 730 CGAAGACATATTTCTGATATATGTCATAAACAATTGTAATTATCATCGTCCACT 789
 |||||
 Db 123 CGAAGACATATTTCTGATATATGTCATAAACAATTGTAATTATCATCGTCCACT 182
 |||||
 QY 790 TGAATCTGTGTTCAATTAAGTCTTAACTCTTCATATTTAGAAATGAGGCTGATGAGTTCC 849
 |||||
 Db 183 TGAATCTGTGTTCAATTAAGTCTTAACTCTTCATATTTAGAAATGAGGCTGATGAGTTCC 242
 |||||
 QY 850 AATTTGAAAGATTTTTCATCACTACTACTGTTTATATGCTTCAAGCCAGAGTGTCTT 909
 |||||
 Db 243 AATTTGAAAGATTTTTCATCACTACTACTGTTTATATGCTTCAAGCCAGAGTGTCTT 302
 |||||
 QY 910 TTTCTATCTACTCTCATACACCAATTAATGCTGAATGTAATTTCAAGGAGATGCGCT 969
 |||||
 Db 303 TTTCTATCTACTCTCATACACCAATTAATGCTGAATGTAATTTCAAGGAGATGCGCT 362
 |||||
 QY 970 AGTGAATTTAACTATGCTGAGCATTCTTGAGTCCAAATATATAAGTATTTGTACTCT 1029
 |||||
 Db 363 AGTGAATTTAACTATGCTGAGCATTCTTGAGTCCAAATATATAAGTATTTGTACTCT 422
 |||||
 QY 1030 TTTGCTGGGTGAGGTGTTCTTTA 1053
 |||||
 Db 423 TTTGCTGGGTGAGGTGTTCTTTA 446
 |||||

RESULT 6 336 bp mRNA linear EST 31-MAR-1995
 R04703
 LOCUS PK27902.r1 Kuwabara Mixed stage C. briggsae Caenorhabditis briggsae
 DEFINITION
 CDNA similar to SP:VE59_LAMB0 P03754 EA59 GENE, mRNA sequence.
 R04703
 ACCESSION
 VERSION R04703.1 GI:754439
 KEYWORDS
 EST

SOURCE
 ORGANISM
 Caenorhabditis briggsae
 Caenorhabditis briggsae
 Burkayota; Metazoa; Chordata; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 336)
 Hillier, L., Chappel, B., Chisoe, S., Clark, N., Couch, J.,
 Dubuque, T., Hawkins, M., Holman, M., Hultman, M., Kuwaba, T.,
 Kuwabara, P., Le, M., Mardis, E., Marra, M., Parsons, D., Rifkin, L.,
 Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Wohlmann, P., and
 Wilson, R.
 TITLE
 JOURNAL
 COMMENT
 Washington University Caenorhabditis briggsae EST project
 Unpublished (1995)
 Other ESTs: pk27902.s1
 Contact: Marra MA
 Washington University Genome Sequencing Center
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1455
 Fax: 314 286 1810
 Email: mmarr@wustl.edu
 PCR F: TGTAACGACGCGCAGGACGAGCAAGTTCACCTGG
 PCR B: CAGGAACAGCTGACCTGATGATGATTTCCGAGGTA
 Source: Washington University Genome Sequencing Center
 PCR amplified DNA is available from Washington University Genome
 Sequencing Center. Aliquots of the library may be requested from P.
 Kuwabara (pkuw@wustl.cam.ac.uk).
 Seq primer: Commercially available M13 reverse dye primer.
 Location/Qualifiers
 1..336
 /organism="Caenorhabditis briggsae"
 /mol_type="mRNA"
 /strain="G16 Gujarat"
 /db_xref="taxon:6238"
 /clone_lib="Kuwabara Mixed stage C. briggsae"
 /note="Vector: Lambda gcl0; Site_1: EcoRI; Site_2: EcoRI;

FEATURES

source
 1..336
 /organism="Caenorhabditis briggsae"
 /mol_type="mRNA"
 /strain="G16 Gujarat"
 /db_xref="taxon:6238"
 /clone_lib="Kuwabara Mixed stage C. briggsae"
 /note="Vector: Lambda gcl0; Site_1: EcoRI; Site_2: EcoRI;

Stage:mixed, Sex:hermaphrodite. Library construction:
 First strand oligo (dT) primed. Second strand was as per
 Gubler/Hoffman. Ligated to EcoRI adaptor. Library is
 non-directional. Library is non-normalized. Library
 constructed by P.R. Kuwabara. Additional details on
 construction of the library are described in P.E.
 Kuwabara and S. Shah, NAR 22: 4414 - 4418 (1994). Adaptor
 sequence: GAATTC CATTGCTGTG

Query Match 13.5%; Score 314; DB 14; Length 336;
 Best Local Similarity 99.4%; Pred. No. 1.8e-62;
 Matches 325; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 972 TGATTTAACTATGCTGAGCATTCTTGAGTCCAAATATAAGTATTTGTACTCTT 1031
 |||||
 Db 336 TGATTTAACTATGCTGAGCATTCTTGAGTCCAAATATAAGTATTTGTACTCTT 277
 |||||
 QY 1032 TGTCTGGTCAAGTGTGTTCTTTAGAGAGATAAAGATCAATGCACTAAAGAACTGA 1091
 |||||
 Db 276 TGTCTGGTCAAGTGTGTTCTTTAGAGAGATAAAGATCAATGCACTAAAGAACTGA 217
 |||||
 QY 1092 AACAGCATGGAATAATCCCTTGGATCTTGACTCGATAGATGATTTTTCAGA 1151
 |||||
 Db 216 AACAGCATGGAATAATCCCTTGGATCTTGACTCGATAGATGATTTTTCAGA 158
 |||||
 QY 1152 GAAAAAATATCATTTGTTTCTGAGTGTGATGATGACCAATCATTCATCAAAATGCT 1211
 |||||
 Db 157 GAAAAAATATCATTTGTTTCTGAGTGTGATGATGACCAATCATTCATCAAAATGCT 96
 |||||
 QY 1212 TGTTTACCAACCCATTCCGCCGATTAAGACATGATGTTGCTGAGGATGATTT 1271
 |||||
 Db 97 TGTTTACCAACCCATTCCGCCGATTAAGACATGATGTTGCTGAGGATGATTT 38
 |||||
 QY 1272 AACGCTACCTCAAAAGATTAAGTTAA 1298
 |||||
 Db 37 AACGCTACCTCAAAAGATTAAGTTAA 11
 |||||

RESULT 7 365 bp mRNA linear EST 10-JAN-2001
 BF737769
 LOCUS CM4-KT0035-131200-610-a12 KT0035 Homo sapiens CDNA, mRNA sequence.
 DEFINITION
 BF737769
 ACCESSION
 VERSION BF737769.1 GI:12064445
 KEYWORDS
 EST
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Burkayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 365)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Britones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bai, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
 O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

TITLE
 JOURNAL
 COMMENT
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 10737800
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL:
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CM4&l2=CM4-KT0035-

FEATURES

131200-610-12&ct3=2000-12-13&ct4=1)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 364.
Location/Qualifiers

1. .365

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_id="KT0035"
/note="Organ: bladder tumor; Vector: puc18; Site: 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 13.3%; Score 309.6; DB 10; Length 365;

Best Local Similarity 98.7%; Pred. No. 1.9e-61; Indels 0; Gaps 0;
Matches 312; Conservative 0; Mismatches 4;

Db 653 AACCTAATGACTTTTCCCAACGACATCTACTATCTTGTGATGTAATTAACAATTG 712
50 AAACCTAATGACTTTTCCCAACGACATCTACTATCTTGTGATGTAATTAACAATTG 109
Qy 713 CAGTCCAGAGCTCATTCGAAAGCAGATATTTTGGATATGTCAATAAACAATTG 772
Db 110 CAGTCCAGAGCTCATTCGAAAGCAGATATTTTGGATATGTCAATAAACAATTG 169
Qy 773 ATTATCATGTCGCACTTGAATCTGTGTCATTAAGCTTTAATTTAGAA 832
Db 170 ATTATCATGTCGCACTTGAATCTGTGTCATTAAGCTTTAATTTAGAA 229
Qy 833 TGAAGCTGATGAGTTCATATTTGAAAAAGTTTCATCACTACTAGTTTGTAGACTT 892
Db 230 TGAAGCTGATGAGTTCATATTTGAAAAAGTTTCATCACTACTAGTTTGTAGACTT 289
Qy 893 CAAGCCGAGTTGCTTTTCTATCTCTGCTCAATCAACAATAAAGCGAAATGAT 952
Db 290 CAAGCCGAGTTGCTTTTCTATCTCTGCTCAATCAACAATAAAGCGAAATGAT 349
Qy 953 CTAAGCGAGATCGCC 968
Db 350 CTAAGCGAGATCGCC 365

RESULT 8
AM984213 262 bp mRNA linear EST 02-JUN-2000
LOCUS PM2-HN0008-120400-003-d07 HN0008 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM984213
VERSION AM984213.1 GI:8175813
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 262)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgensen,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20200263
PubMed 10737800
COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FADESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ct3=PM2-HN0008-120
400-003-d07&ct3=2000-04-12&ct4=1)

Seq primer: puc 18 forward
High quality sequence start: 13
High quality sequence stop: 218.
Location/Qualifiers

FEATURES

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_id="HN0008"
/note="Organ: head normal; Vector: puc18; Site 1: SmaI;
Site 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

ORIGIN

Query Match 10.3%; Score 238.6; DB 10; Length 262;

Best Local Similarity 98.1%; Pred. No. 6.7e-45; Indels 1; Gaps 1;
Matches 252; Conservative 0; Mismatches 4;

Qy 1311 GGGAGCAGCTTTTCTAATTAAGTGAAGTGAATCTGACATTTGCGCAACCAATTAA 1370
Db 262 GGGAGCAGCTTTTCTAATTAAGTGAAGTGAATCTGACATTTGCGCAACCAATTAA 203
Qy 1371 CACACGTCGCACTGTCATGAATTTCTGAAAGATTACCCCTCTAATGATGAGGT 1430
Db 202 CACACGTCGCACTGTCATGAATTTCTGAAAGATTACCCCTCTAATGATGAGGT 143
Qy 1431 AAGGAGCTTTCAATTTCAATGTCGGTATCGATTGGCCATCTAATCTGAT 1430
Db 142 AAGGAGCTTTCAATTTCAATGTCGGTATCGATTGGCCATCTAATCTGAT 83
Qy 1491 AGCTTTAAGAGGTATGTTAAACCATGCTTAATTTGCTGAGATTAACTAGTGC 1550
Db 82 AGCTTTAAGAGGTATGTTAAACCATGCTTAATTTGCTGAGATTAACTAGTGC 24
Qy 1551 AATGCTTCACTAAGG 1567
Db 23 AATGCTTCACTAAGG 7

RESULT 9
BF773183 254 bp mRNA linear EST 12-JAN-2001
LOCUS CM4-IT0046-151200-606-b11 IT0046 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF773183
VERSION BF773183.1 GI:12121083
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 254)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jorgensen,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=CM4&c2=CM4-IT0046-151200-606-b1&c3=2000-12-15&c4=1)
Seq primer: puc 18 forward
High quality sequence start: 7
High quality sequence stop: 253.
Location/Qualifiers

FEATURES

source

1..254
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="IT0046"
/note="Organ: epid. tumor; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 8.9%; Score 205.8; DB 10; Length 254;
Best Local Similarity 98.0%; Pred. No. 3e-37; 2; Indels 3; Gaps 3;
Matches 240; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 617 AAGGTGATGACAGATGAACCTGCTTCATCGAATAAACAATGACCTTTGGCCACGA 676
DB 11 AAGGTGATGACAGATGAACCTTTGG-TCATCGAATCAATGACCTTTGGCCACGA 69
QY 677 CATCTACTATCT-TGTGATAGTAATAAACAATGACCTGTCAGAGCTCATTTGGAAC 735
DB 70 CATCTACTATCTCTGTGATAGTAATAAACAATGACCTGTCAGAGCTCATTTGGAAC 129
QY 736 AGATATTTCTGATATTTGT-CATTAACAATTTAGTAATTATCATGTCACCTTGAA 794
DB 130 AATATTTCTGATATTTGTCTCAATAAACAATTTAGTAATTATCATGTCACCTTGAA 189
QY 795 CTGTGATCATTAAGCTTTCACTTTCAATTTAGTAAGAGCTGATGATTCATATT 854
DB 190 CTGTGATCATTAAGCTTTCACTTTCAATTTAGTAAGAGCTGATGATTCATATT 249
QY 855 TGAAG 859
DB 250 TGAAG 254

RESULT 10

BF348952 206 bp mRNA linear EST 22-NOV-2000
LOCUS MRL-DT0059-110700-002-g08 DT0059 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF348952
VERSION BF348952.1 GI:11308026
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 206)

AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.G.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

PUBMED 10737800

COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MRL&c2=MRL-DT0059-110700-002-g08&c3=2000-07-11&c4=1)
Seq primer: puc 18 forward
High quality sequence stop: 206.
Location/Qualifiers

FEATURES

source

1..206
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="DT0059"
/note="Organ: dentis. drash; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 8.7%; Score 202.8; DB 10; Length 206;
Best Local Similarity 99.0%; Pred. No. 1.5e-36;
Matches 204; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1410 CCCCTAAGTATGAGTGTGTAAGACGCTTCATTTCAATGTCGCTAATCATTTGG 1469
DB 1 CCCCTAAGTATGAGTGTGTAAGACGCTTCATTTCAATGTCGCTAATCATTTGG 60
QY 1470 CCATACCTAATCTCGAATAGCTTTAAGAAAGTTATGTTAAACCATGCTTAATT 1529
DB 61 CCATACCTAATCTCGAATAGCTTTAAGAAAGTTATGTTAAACCATGCTTAATT 120
QY 1530 GGTGATTAATCATGTGTGTAAGCTTCCTCAAGAAAAAATTTGAGGAGTTG 1589
DB 121 GGTGATTAATCATGTGTGTAAGCTTCCTCAAGAAAAAATTTGAGGAGTTG 180
QY 1590 ACTGAATTTTATCTATTAAATGAAT 1615
DB 181 ACTGAATTTTATCTATTAAATGAAT 206

RESULT 11

CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
LOCUS CNS0039G/c
DEFINITION Drosophila melanogaster genome survey sequence TERN end of BAC # BACH08K10 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster

Best Local Similarity 22.2%; Pred. No. 0.0018;
 Matches 141; Conservative 204; Mismatches 289; Indels 2; Gaps 1;

QY 1569 AAAAAAATTCGGGGAGTGGAGTATTTTATTCATTAATCAATGAGTCTACTTC 1638
 Db 315 AAAAAAATTCGGGGAGTGGAGTATTTTATTCATTAATCAATGAGTCTACTTC 374
 QY 1629 TTTCTTTGACCTTAACAACCAATTTTACATTTCCGATATCCGATTTTCCAGTCTC 1688
 Db 375 TTTCTTTGACCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 434
 QY 1689 ATCAAAAGACGTAAAGATTAACATTTTAAACAAGAAATAGTCAATCCAGTCTCTG 1748
 Db 435 TTTCAACCCCAAHNHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHAHA 494
 QY 1749 TAGGAATGCTTTATTTTCTCTGAGGAATATACCGGCTCTCTCAATTAACATAA 1808
 Db 495 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 554
 QY 1809 CTCACATATATAGTAACTTATTTATTTATTAATAAATTAATTTTGGCGGCAACA 1868
 Db 555 YAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 614
 QY 1869 CAGATCTCTCTTTAAGTATCTCTATTAATAGTCTTTTCCATCTTAATAATAGT 1928
 Db 615 AAAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 674
 QY 1929 ATGAATACGGGGAGTGGATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1988
 Db 675 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 734
 QY 1989 GATTAACCACTGTATTCATGTGATGATGATGATGATGATGATGATGATGATGAT 2048
 Db 735 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 794
 QY 2049 AATGATATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 2108
 Db 795 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 852
 QY 2109 TACTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2168
 Db 853 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 912
 QY 2169 CGCT 2204
 Db 913 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 948

RESULT 15
 Locus008H1
 DEFINITION
 Drosophila melanogaster genome survey sequence TERT3 end of BAC #
 BACR17001 of RPCI-98 library from Drosophila melanogaster (fruit
 fly) genomic survey sequence.
 ACCESSION
 AL051882
 VERSION
 AL051882.1 GI:4933734
 SOURCE
 Drosophila melanogaster (fruit fly)
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 961)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Ooeogawa and
 Aaron Mammocier in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
 1..961
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR17001"
 /clone_1lb="RPCI-98"
 /note="Tend : TERT3"

ORIGIN

Query Match 2.6%; Score 59.8; DB 29; Length 961;
 Best Local Similarity 39.3%; Pred. No. 0.0034;
 Matches 175; Conservative 39; Mismatches 231; Indels 0; Gaps 0;

QY 1485 CTGAATAGCTTTAAGAGGTTATGTTTAAACATGCTTAATTTGCTGATTAACATA 1544
 Db 137 CTGATATATTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 196
 QY 1545 GTAGTCAATGCTTACCTTAAGAAAAAATTTTCAAGGAGTGTGATGATTTTATC 1604
 Db 197 TCTTAAATGAAAAATCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 256
 QY 1605 TATTAATGAATTAAGTCTTCTTTTAAACCTTAACCAATTTTAAATTTTCC 1664
 Db 257 TTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 316
 QY 1665 GATATGCAATTTTACCAATGCTTCAAGACAGTAAGTAACATTTGAACAAAGA 1724
 Db 317 TTTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 376
 QY 1725 ATAGTCAATGCAATGCTTCTGTAAGAGGCTTTTCTCTGAGGATATA 1784
 Db 377 ATTTCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 436
 QY 1785 CCGGCTCTTTCAATTAACATACTCAACATATAGTAACCTTAATTTTAAATA 1844
 Db 437 ATCAATCAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATA 496
 QY 1845 ACCGCAATTTATTTGGGCAACAGAGATCTCTTTAAGTACCTCTATTAATAC 1904
 Db 497 ATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 556
 QY 1905 GTTTTCATCTAAATTTAATTAAGTA 1929
 Db 557 TTATTAATTAAGAAAAAATGCTGCTA 581

Search completed: October 6, 2004, 09:36:45
 Job time : 5958 secs

CC adjacent genes in a host such as bacteria, fungus, yeast, animal, mammal
 CC or plant cell (e.g. apple, Arabidopsis, bajra, banana, barley, bean,
 CC beet, blackgram, chickpea, chili, cucumber, eggplant, fabbean, maize,
 CC melon, millet, oat,okra, papaya, peanut, pea, pepper, pigeonpea,
 CC pineapple, pumpkin, rice, squash, sugarcane, sunflower, sweet potato,
 CC tea, tomato, tobacco, watermelon or wheat). to improve expression of
 CC multiple gene expression constructs in both prokaryotic and eukaryotic
 CC host cells. Gene expression in transgenic plant host cells or transgenic
 CC plants is useful for producing various desirable traits or gene products,
 CC including control of flowering, herbicide resistance, fungicide
 CC resistance, insect resistance, plant size or form, nutrient content,
 CC drought-tolerance, various pathway intermediates for the modulation of
 CC pathways already existing in the host for the synthesis of new products,
 CC and examination and determination of ectopic gene expression and gene
 CC function, particularly as related to the fields of genomics and
 CC proteomics. The present sequence represents a bacteriophage lambda spacer
 CC polynucleotide used in the method of the invention

XX Sequence 2322 BP; 694 A; 446 C; 415 G; 767 T; 0 U; 0 Other;

Query Match 100.0%; Score 2322; DB 7; Length 2322;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTTTCTAATTAACTTTGTGAGGTTACCACTAAGGTTGAGGCTCAAGAG 60
 DB 1 AAGCTTTTCTAATTAACTTTGTGAGGTTACCACTAAGGTTGAGGCTCAAGAG 60
 QY 61 GTGTGCTGTGCTGAGTAAATACTGAGCTGAGCTTAATTTCTATATGTTGTC 120
 DB 61 GTGTGCTGTGCTGAGTAAATACTGAGCTGAGCTTAATTTCTATATGTTGTC 120
 QY 121 TTTCTGCAAAAAGTGGGGAAGTGAATGAATTAATTTCTAATTTCTGATCAT 180
 DB 121 TTTCTGCAAAAAGTGGGGAAGTGAATGAATTAATTTCTAATTTCTGATCAT 180
 QY 181 ACCCTCGAGCATTTAAGCATTTGCTTAAGATTGCTGCGAAGAGGATTTTTT 240
 DB 181 ACCCTCGAGCATTTAAGCATTTGCTTAAGATTGCTGCGAAGAGGATTTTTT 240
 QY 241 CATTTGACTTACCTGATCTGTGATTAATCAATGCTTTTAAAGGTTCCAGCTCT 300
 DB 241 CATTTGACTTACCTGATCTGTGATTAATCAATGCTTTTAAAGGTTCCAGCTCT 300
 QY 301 AATCTTACTGACCATTAATATTTTGAATGTTTCAATAGAAGCTGATCAACG 360
 DB 301 AATCTTACTGACCATTAATATTTTGAATGTTTCAATAGAAGCTGATCAACG 360
 QY 361 GACTGCGATTAATAAGTGTGATCCGAATTTGCTCACTTCAAGTAAACACCTCAGA 420
 DB 361 GACTGCGATTAATAAGTGTGATCCGAATTTGCTCACTTCAAGTAAACACCTCAGA 420
 QY 421 GTTAAACACCTTAAGTCTACACGGAATGTCGAATCCGAGAGGATTAATTAATGCT 480
 DB 421 GTTAAACACCTTAAGTCTACACGGAATGTCGAATCCGAGAGGATTAATTAATGCT 480
 QY 481 TCTCTTACCGTAGAGCTTTTCCACATGACAGATTTTGAACCTTTGACGACTACTG 540
 DB 481 TCTCTTACCGTAGAGCTTTTCCACATGACAGATTTTGAACCTTTGACGACTACTG 540
 QY 541 GAATGAGTGTGATTAATTTGCTACACCATTCGTCGATCGAGTAAGTGTAAAGTTGCT 600
 DB 541 GAATGAGTGTGATTAATTTGCTACACCATTCGTCGATCGAGTAAGTGTAAAGTTGCT 600
 QY 601 AAAAAACAGAGAGCAAGGAGTGAAGTGAACCTCTGGTTACGTAATTAACCTAAT 660
 DB 601 AAAAAACAGAGAGCAAGGAGTGAAGTGAACCTCTGGTTACGTAATTAACCTAAT 660
 QY 661 GACTTTTGGCAAGCATCTAATCTTGTGATAGTAATAAACAATTTGATGCA 720
 DB 661 GACTTTTGGCAAGCATCTAATCTTGTGATAGTAATAAACAATTTGATGCA 720
 QY 721 GAGCTAATCGAAGCAGATTTTCTGATATGTCATPAAACAATTTAGTAATTAATCA 780

DB 721 GAGCTAATCGAAGCAGATTTTCTGATATGTCATPAAACAATTTAGTAATTAATCA 780
 QY 781 TCGTCCACTGAATCTGTGATCAATACGCTTAACCTTCATATTTAGAAATGAGCTG 840
 DB 781 TCGTCCACTGAATCTGTGATCAATACGCTTAACCTTCATATTTAGAAATGAGCTG 840
 QY 841 ATGAGTCCATTTTGAAGTTTCACTACTACTTATGTTTGTGATAGCTCAAGCAG 900
 DB 841 ATGAGTCCATTTTGAAGTTTCACTACTACTTATGTTTGTGATAGCTCAAGCAG 900
 QY 901 AGTTGTCTTTTCTATCTACTCTTCATCAACCAATTAATGCTGAATGAAATTTCAAGCG 960
 DB 901 AGTTGTCTTTTCTATCTACTCTTCATCAACCAATTAATGCTGAATGAAATTTCAAGCG 960
 QY 961 AGATGCGCTGATTTTAATTAATGCTGCGAGCATCTTGGTGCATTAATAAGAT 1020
 DB 961 AGATGCGCTGATTTTAATTAATGCTGCGAGCATCTTGGTGCATTAATAAGAT 1020
 QY 1021 TGTGTACCTTTTGTGAGGTCAGGTTGTTCTTTAGAGAGTAAAGATCAATGCACTA 1080
 DB 1021 TGTGTACCTTTTGTGAGGTCAGGTTGTTCTTTAGAGAGTAAAGATCAATGCACTA 1080
 QY 1081 AAGGAACTGAAACAGCGATCGAAATATCCCTTTGGGATTTCTGACTGATTAAGTCTA 1140
 DB 1081 AAGGAACTGAAACAGCGATCGAAATATCCCTTTGGGATTTCTGACTGATTAAGTCTA 1140
 QY 1141 TTAATTTTCAAGAGAAAATATTCATTTGCTGGGTGGGATTTGACCAATCATTTCCA 1200
 DB 1141 TTAATTTTCAAGAGAAAATATTCATTTGCTGGGTGGGATTTGACCAATCATTTCCA 1200
 QY 1201 TTCAAAATTTGTTTGTACCAACCCATTCGCGCCGATAAAGCATGAATGTTGTCGTG 1260
 DB 1201 TTCAAAATTTGTTGTTTGTACCAACCCATTCGCGCCGATAAAGCATGAATGTTGTCGTG 1260
 QY 1261 GGCATAGAAATTAACCGTCACTCAAAAAGTATGTTAAATCACTGAATCCGGGAGCACTT 1320
 DB 1261 GGCATAGAAATTAACCGTCACTCAAAAAGTATGTTAAATCACTGAATCCGGGAGCACTT 1320
 QY 1321 TTTCTAATTAATGAAGTGAATCTGACATTTGCGCAACCATTTAACAACAGTGGC 1380
 DB 1321 TTTCTAATTAATGAAGTGAATCTGACATTTGCGCAACCATTTAACAACAGTGGC 1380
 QY 1381 AACTGTCAATGAATTTCTGAAAGATTAACCTCTTAAGTATGAGTGTAAAGACGCTT 1440
 DB 1381 AACTGTCAATGAATTTCTGAAAGATTAACCTCTTAAGTATGAGTGTAAAGACGCTT 1440
 QY 1441 TCAATTTCAATGTCGGTATTCGATTTGGCCATTAATCTAATCTGATATGCTTTAAGA 1500
 DB 1441 TCAATTTCAATGTCGGTATTCGATTTGGCCATTAATCTAATCTGATATGCTTTAAGA 1500
 QY 1501 AGGTTATGTTAAACCATGCTTAATTTGCTGAGATTAACATGATGTCATGCTTTCA 1560
 DB 1501 AGGTTATGTTAAACCATGCTTAATTTGCTGAGATTAACATGATGTCATGCTTTCA 1560
 QY 1561 CTTAAGGAAAAAAACATTTGAGGAGTTGACTGAATTTTATCTATTAATGAATAAGTG 1620
 DB 1561 CTTAAGGAAAAAAACATTTGAGGAGTTGACTGAATTTTATCTATTAATGAATAAGTG 1620
 QY 1621 CTTACTCTCTCTTTTGAACCTCAACAAACCAATTTTCAATTCGATATGCAATTTTCA 1680
 DB 1621 CTTACTCTCTCTTTTGAACCTCAACAAACCAATTTTCAATTCGATATGCAATTTTCA 1680
 QY 1681 CCAATGCTCATCAAGACATGAATTAACATTTGAACAAGGAATGTCATTTCAACCA 1740
 DB 1681 CCAATGCTCATCAAGACATGAATTAACATTTGAACAAGGAATGTCATTTCAACCA 1740
 QY 1741 TCTGTCTGTAAGGAATGCTTAATTTTCTGATGCAAGAAATATACCGCTCTTTCAATA 1800
 DB 1741 TCTGTCTGTAAGGAATGCTTAATTTTCTGATGCAAGAAATATACCGCTCTTTCAATA 1800
 QY 1801 ACACTTAACCTCAACATATGTAACCTTAATTTATTAATAACCGCAATTTAATTTG 1860

```

Db      1801 AACCTAACTCCGACATATAGTAACCCCTAATTTTATTAATAAACCGCAATTTATTTG 1860
Qy      1861 CGGCAACACAGAGATCTCTCTTTAAAGTAATCTCTAATTAACATGCTTTTCCATTAATAA 1920
Db      1861 CGGCAACACAGAGATCTCTCTTTAAAGTAATCTCTAATTAACATGCTTTTCCATTAATAA 1920
Qy      1921 TTAGTAGTATTTAGACTTAACGGGGCATGCTATTTGATTTTCCATTTAGCTTTCTGCT 1980
Db      1921 TTAGTAGTATTTAGACTTAACGGGGCATGCTATTTGATTTTCCATTTAGCTTTCTGCT 1980
Qy      1981 TCCCTTTGGATTAACCACTGTTATTCATGTTGATGATGACGCTGTTATTAACCAAGATAT 2040
Db      1981 TCCCTTTGGATTAACCACTGTTATTCATGTTGATGATGACGCTGTTATTAACCAAGATAT 2040
Qy      2041 AGCTATTAATATCATATATATAGTATGCGGACGATAGCTCTTCAGGCTTCGAAAGAAC 2100
Db      2041 AGCTATTAATATCATATATATAGTATGCGGACGATAGCTCTTCAGGCTTCGAAAGAAC 2100
Qy      2101 GTTTCAAGACTAATTAAGCCGATAGATAGCCAGACTTCTGATGCTTTTCAATAGTG 2160
Db      2101 GTTTCAAGACTAATTAAGCCGATAGATAGCCAGACTTCTGATGCTTTTCAATAGTG 2160
Qy      2161 TTAACCTTCGCTCGCTCGCTCATTAACAGACATTCATCAGTATAGCGGAAAGTATGCA 2220
Db      2161 TTAACCTTCGCTCGCTCGCTCATTAACAGACATTCATCAGTATAGCGGAAAGTATGCA 2220
Qy      2221 TCGCTGGGTGGGGAAGTGTGTAAGAAAGAAAGAGTGCCTGTTTGAATCACTGC 2280
Db      2221 TCGCTGGGTGGGGAAGTGTGTAAGAAAGAAAGAGTGCCTGTTTGAATCACTGC 2280
Qy      2281 TATCTTCTTACTGTTATGACAGTGTGTATGCGGTGGACACA 2322
Db      2281 TATCTTCTTACTGTTATGACAGTGTGTATGCGGTGGACACA 2322

RESULT 2
ID      ACA55352
ID      ACA55352 standard, DNA; 9980 BP.
XX
XX      ACA55352;
XX
XX      06-JUN-2003 (first entry)
XX
XX      pIAO-P/L plasmid series, pIAO-P/L-lambda nucleotide sequence.
XX
XX      PiggyBac; transposon; eukaryotic transformation vector; ds;
XX      circular.
XX      circular.
XX
XX      Synthetic.
XX
XX      US2002173634-A1.
XX
XX      21-NOV-2002.
XX
XX      30-OCT-2001; 2001US-00001189.
XX
XX      31-OCT-2000; 2000US-0244677P.
XX      01-NOV-2000; 2000US-0244984P.
XX
XX      (FRAS/) FRASER M J.
XX      (LIXX/) LI X.
XX      (BEAM/) BEAM T.
XX      (HUA/) HUA-VAN A.
XX
XX      Fraser MJ, Li X, Beam T, Hua-Van A;
XX
XX      WPI: 2003-352597/33.
XX      P-PSDB; AB070356.
XX
XX      New DNA molecule in the transposon piggyBac, useful for transferring
XX      PT genes into host cells or embryos for transforming the cells of embryos
XX      PT that can be used in making transgenic organisms.

```

```

XX      Example 2; Fig 2(C2); 151bp; English.
PS      The invention describes a DNA molecule comprising at least 163
XX      consecutive nucleotide base pairs of the 3' terminal region beginning at
CC      the 3' terminal base pair, and at least 125 consecutive nucleotide base
CC      pairs of the 5' terminal region beginning at the 5' terminal base pair, of
CC      the piggyBac molecule. The region extending from the restriction site SacI
CC      to the end of the piggyBac molecule. The DNA molecule in the transposon
CC      piggyBac is useful for transferring genes into host cells or embryos for
CC      transforming the cells of embryos. The transformed cells or embryos are
CC      useful for developing or making transgenic organisms. This sequence
CC      represents plasmid used in the creation of minimal sequence eukaryotic
CC      transformation vector piggyBac
SQ
Sequence 9980 BP; 2697 A; 2342 C; 2139 G; 2801 T; 0 U; 1 Other;

```

```

Query Match      99.9%; Score 2319.4; DB 7; Length 9980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2320; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      2 AGCTTTCTAATTTAACTTTGTGACGTTACCACTAAGGTTAGGCTCAAGAGG 61
Db      6608 AGCTTTCTAATTTAACTTTGTGACGTTACCACTAAGGTTAGGCTCAAGAGG 6667
Qy      62 TGTGTCTGTGCTAGTAAATTAAGTACCTGTGAGCTTAATTTCTATTTGTTCT 121
Db      6668 TGTGTCTGTGCTAGTAAATTAAGTACCTGTGAGCTTAATTTCTATTTGTTCT 6727
Qy      122 TTTCTGCAAAAAAGTGGGGAAGTGAATGAATTTATTTCTAATTTATCTGATCATA 181
Db      6728 TTTCTGCAAAAAAGTGGGGAAGTGAATGAATTTATTTCTAATTTATCTGATCATA 6787
Qy      182 CCTCCGAGCATTTTAAAGCATTTGCTAATAGTTCTGCTGGAAGAGTATTTTTC 241
Db      6788 CCTCCGAGCATTTTAAAGCATTTGCTAATAGTTCTGCTGGAAGAGTATTTTTC 6847
Qy      242 ATTGTACTTTACCTTCATCTCTGTTCAATATCATGCTTTTAAACGTTGACCTTCTA 301
Db      6848 ATTGTACTTTACCTTCATCTCTGTTCAATATCATGCTTTTAAACGTTGACCTTCTA 6907
Qy      302 ATCCATATCGACATTAATTAATTTTAAAGTGTTCATTAAGAAAGCTGTGAATCAAGG 361
Db      6908 ATCCATATCGACATTAATTAATTTTAAAGTGTTCATTAAGAAAGCTGTGAATCAAGG 6967
Qy      362 ACTGGAATATATAGTGTGTATCCAGATTTGTCACTCAAGTAAACCACTCAAGAG 421
Db      6968 ACTGGAATATATAGTGTGTATCCAGATTTGTCACTCAAGTAAACCACTCAAGAG 7027
Qy      422 TTTAAACACCTTAAGTTCACCGAATGTCTCAATATCCGAGCGATATATTTTGTCT 481
Db      7028 TTTAAACACCTTAAGTTCACCGAATGTCTCAATATCCGAGCGATATATTTTGTCT 7087
Qy      482 CTCTGACCGTGAAGCTTTCCACATGACAGATTTTGGAACTCTGACGATCACTACGCGG 541
Db      7088 CTCTGACCGTGAAGCTTTCCACATGACAGATTTTGGAACTCTGACGATCACTACGCGG 7147
Qy      542 AATGAGTTCATTTATTTGCTACACCATGCGTCACTGAGTGAAGTGCCTTAATGTTCTGA 601
Db      7148 AATGAGTTCATTTATTTGCTACACCATGCGTCACTGAGTGAAGTGCCTTAATGTTCTGA 7207
Qy      602 AAAAAAGAGAGCAAGAGTGTGATGAGTGAACCTGTGTTCAATGAATTAACCTAATG 661
Db      7208 AAAAAAGAGAGCAAGAGTGTGATGAGTGAACCTGTGTTCAATGAATTAACCTAATG 7267
Qy      662 ACTTTTGGCAACGACATCTAATCTCTGTGATAGTAAATTAACCAATTTGATGCGAG 721
Db      7268 ACTTTTGGCAACGACATCTAATCTCTGTGATAGTAAATTAACCAATTTGATGCGAG 7327
Qy      722 AGCTCATTCGAAGAGATATTTCTGATATGCTCAATTAACCAATTTAGTGAATTTATCAT 781
Db      7328 AGCTCATTCGAAGAGATATTTCTGATATGCTCAATTAACCAATTTAGTGAATTTATCAT 7887

```

QY 782 CGTCCACTGATCTGCTGCTCATTAAGCTCTTAACCTCTCATATTTAGAAATGAGCGCTGA 841
 DB 7388 CGTCCACTGATCTGCTGCTCATTAAGCTCTTAACCTCTCATATTTAGAAATGAGCGCTGA 7447
 QY 842 TGAATCTCATATTTGAAAAGTTTTCATCACTACTTAACTTTTGTATGCTTGAAGCCAGA 901
 DB 7448 TGAATCTCATATTTGAAAAGTTTTCATCACTACTTAACTTTTGTATGCTTGAAGCCAGA 7507
 QY 902 GTGTCTTTTCTATCTACTCTCATACCAACCAATTAATGCTGAATATATTTCTAAGCGGA 961
 DB 7508 GTGTCTTTTCTATCTACTCTCATACCAACCAATTAATGCTGAATATATTTCTAAGCGGA 7567
 QY 962 GATGCGCTAGTATTTTAACTATTTGCTGCGAGATTTCTGATGCCAATATATAAGTATT 1021
 DB 7568 GATGCGCTAGTATTTTAACTATTTGCTGCGAGATTTCTGATGCCAATATATAAGTATT 7627
 QY 1022 GTGTACCTTTTGTGCTGGGCTGCTGCTTGTAGGAGAGTAAAGATTCAAATGCACTAA 1081
 DB 7628 GTGTACCTTTTGTGCTGGGCTGCTGCTTGTAGGAGAGTAAAGATTCAAATGCACTAA 7687
 QY 1082 ACGAACTGAAACCAAGGATCGAAATATATCCCTTTGGGATTTCTGACTCGATTAAGTCTAT 1141
 DB 7688 ACGAACTGAAACCAAGGATCGAAATATATCCCTTTGGGATTTCTGACTCGATTAAGTCTAT 7747
 QY 1142 TATTTTCAAGAAAATAATTCATTGTTTCTGCGTTGCTGATTCGACCAATCACTTCAT 1201
 DB 7748 TATTTTCAAGAAAATAATTCATTGTTTCTGCGTTGCTGATTCGACCAATCACTTCAT 7807
 QY 1202 TCAAAATTTGTTGTTTCCACACACCATTTCGCGCCGATTAAGATGTTGCTGCTGG 1261
 DB 7808 TCAAAATTTGTTGTTTCCACACACCATTTCGCGCCGATTAAGATGTTGCTGCTGG 7867
 QY 1262 GCATAGATTAATTAACCGTCACTCAAAAGTATATTAATCACTGAATCCGGAGCACTTT 1321
 DB 7868 GCATAGATTAATTAACCGTCACTCAAAAGTATATTAATCACTGAATCCGGAGCACTTT 7927
 QY 1322 TTCTATTAATGAAAAGTGAATCTGCAATTTCTGCGAACCACTTTAACAACAGTCCGA 1381
 DB 7928 TTCTATTAATGAAAAGTGAATCTGCAATTTCTGCGAACCACTTTAACAACAGTCCGA 7987
 QY 1382 ACTGTCATGATTAATTCGAAAAGTTAACCCCTCTAATGAATGAGTGTGAAGCGCTT 1441
 DB 7988 ACTGTCATGATTAATTCGAAAAGTTAACCCCTCTAATGAATGAGTGTGAAGCGCTT 8047
 QY 1442 CATTTTCATGTCGGCTAATCGATTTGGCCATCTACTTAATCTCGAATGCTTTAAGAA 1501
 DB 8048 CATTTTCATGTCGGCTAATCGATTTGGCCATCTACTTAATCTCGAATGCTTTAAGAA 8107
 QY 1502 GGTATATGTTTAAACCATCGCTTAATTTGCTGAATTAACATAGTCAATGCTTTAC 1561
 DB 8108 GGTATATGTTTAAACCATCGCTTAATTTGCTGAATTAACATAGTCAATGCTTTAC 8167
 QY 1562 CTAAAGAAAATAATCTTACAGGAGTGAAGTGAATTTTATCTAATTAATGAATAGTGC 1621
 DB 8168 CTAAAGAAAATAATCTTACAGGAGTGAAGTGAATTTTATCTAATTAATGAATAGTGC 8227
 QY 1622 TTAATCTCTCTTTTGAAGCTTAACAAACCAATTTTAACTTTCCGATTTTCAC 1681
 DB 8228 TTAATCTCTCTTTTGAAGCTTAACAAACCAATTTTAACTTTCCGATTTTCAC 8287
 QY 1682 CATGCTCATCAAGACAGTAAAGATTAACATTTGTAACAAAGAAATAGTCAATTCACCAT 1741
 DB 8288 CATGCTCATCAAGACAGTAAAGATTAACATTTGTAACAAAGAAATAGTCAATTCACCAT 8347
 QY 1742 CTGCTCGAAGAAAGCTTATTTTCTACTGAGGAATATACCCGCTCTTCAATTA 1801
 DB 8348 CTGCTCGAAGAAAGCTTATTTTCTACTGAGGAATATACCCGCTCTTCAATTA 8407
 QY 1802 CACTAAATCTCCAAATATAGTAACCTTAATTTTATTAATAACCGCAATTTATTTGGC 1861
 DB 8408 CACTAAATCTCCAAATATAGTAACCTTAATTTTATTAATAACCGCAATTTATTTGGC 8467
 QY 1862 GGCAACACAGAGATCTCTTTTAAGTTAAGTCTCTATTAATACATGCTTTCCATCTAATAAT 1921

DB 8468 GGCAACACAGAGATCTCTTTTAAGTACTCTATTAACATAGCTTTCCATCAATAAAT 8527
 QY 1922 TAGTATGATTAAGATCTTAACCGGGCATGTAATGATTTCCAAATTAAGCTTCTGCT 1981
 DB 8528 TAGTATGATTAAGATCTTAACCGGGCATGTAATGATTTCCAAATTAAGCTTCTGCT 8587
 QY 1982 CCTTTTGAATTAACCATGTTATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 2041
 DB 8588 CCTTTTGAATTAACCATGTTATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 8647
 QY 2042 GTCTATTAATGATTAATAGTATTCGCGAAGATTAAGTCTCTTCAAGGCTTCAAGAAAGCG 2101
 DB 8648 GTCTATTAATGATTAATAGTATTCGCGAAGATTAAGTCTCTTCAAGGCTTCAAGAAAGCG 8707
 QY 2102 TTTCAAGTACTAATTAACCGCATATGATGATGATGATGATGATGATGATGATGATGATGAT 2161
 DB 8708 TTTCAAGTACTAATTAACCGCATATGATGATGATGATGATGATGATGATGATGATGATGAT 8767
 QY 2162 TAACTTCGCTCTCTGCTCATTAACGACATTCATCAATTAATGATGATGATGATGATGAT 2221
 DB 8768 TAACTTCGCTCTCTGCTCATTAACGACATTCATCAATTAATGATGATGATGATGATGAT 8827
 QY 2222 GCTGGGTGTTGGGGAAGTCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2281
 DB 8828 GCTGGGTGTTGGGGAAGTCTGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 8887
 QY 2282 ATCTTCTACTGTTATGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2342
 DB 8888 ATCTTCTACTGTTATGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 8942

RESULT 3
 ABQ80306
 ID ABQ80306 standard; DNA, 1469 BP.
 XX
 AC ABQ80306;
 XX
 DT 27-JUN-2003 (first entry)
 XX
 DE Lambda fragment in plasmid DNA.
 XX
 KW Primer; PCR; amplify; lambda phage; primer material; insertion; SNP;
 KW deletion; single nucleotide polymorphism; sequencing; diagnosis; ss.
 OS Synthetic.
 XX
 FN W02003027991-A1.
 XX
 PD 03-APR-2003.
 XX
 PF 24-SEP-2002; 2002MO-JP009766.
 XX
 PR 25-SEP-2001; 2001JP-00291249.
 XX
 PA (DNAP-) DNAPFORM KK.
 PA (RIKE) RIKEN KK.
 PI Hayashizaki Y;
 XX
 DR WPI: 2003-354676/33.
 XX
 PT Printed material useful as a delivery and storage system for oligomer
 PT and/or polymer, comprises a support having an oligomer and/or polymer
 PT applied on it.
 XX
 PS Example 1; Page 31-32; 91PP; English.
 XX
 CC The sequences given in ABQ80304-05 are primers which were used to amplify
 CC DNA for use in the material of the invention. The primers amplify a 1377
 CC bp portion of lambda phage DNA contained in a plasmid. The amplified
 CC sequence (ABQ80306) was attached to the primer material of the invention
 CC which comprises at least one support having at least one oligomer and/or


```
Db      392 GTTGCAATGCGCACTGTTATACCAACGATATAGTATTAAAGCATATATAGATCGCC 451
QY      2069 GAACGATTAGCTCTTCAGGCTTCTGAAAGAGCGTTTCAATGACTAATAAAGCCATGATA 2128
Db      452 GAACGATTAGCTCTTCAGGCTTCTGAAAGAGCGTTTCAATGACTAATAAAGCCATGATA 511
QY      2129 GCCACGAGCTCTGAGCCATTTTTCATAGTGTAACTTCGCTCGCTCCTGCTCAATACAGA 2188
Db      512 GCCACGAGCTCTGAGCCATTTTTCATAGTGTAACTTCGCTCGCTCCTGCTCAATACAGA 571
QY      2189 CATTCACTACAGTTATGCGGAAAGGATATGCACTGCGGTGCGGGAAGTCGTGAAGAA 2248
Db      572 CATTCACTACAGTTATGCGGAAAGGATATGCACTGCGGTGCGGGAAGTCGTGAAGAA 631
QY      2249 AAGAAAGTCACTGCGTGTGATGACATGCTATCTTCTTACGCTGTTATGCGAGTGTGA 2308
Db      632 AAGAAAGTCACTGCGTGTGATGACATGCTATCTTCTTACGCTGTTATGCGAGTGTGA 691
QY      2309 GTGGGTGGCACACA 2322
Db      692 GTGGGTGGCACACA 705

RESULT 5
ACD19000 standard; DNA; 10523 BP.
XX      ACD19000;
AC      ACD19000;
XX      27-OCT-2003 (revised)
DT      21-AUG-2003 (first entry)
XX      E. coli 0157 unique DNA sequence OZ1D_13.
DE      OZ1D; ds; acute haemorrhagic colitis; haemolytic uraemic syndrome;
KM      food poisoning.
XX      Escherichia coli; strain 0157:H7.
XX      US2003023075-A1.
XX      30-JAN-2003.
XX      01-APR-2002; 2002US-00114170.
XX      04-DEC-1998; 98US-0110955P.
XX      03-DEC-1999; 99US-00453702.
XX      (BLAT/) BLATTNER F. R.
XX      (BURL/) BURLAND V D.
XX      (PERN/) PERNA N T.
XX      (PLUNK/) PLUNKETT G.
XX      (WELC/) WELCH R.
XX      Biatner FR, Burland VD, Perna NT, Plunkett G, Welch R;
PI      WPI; 2003-479497/45.
XX      New DNA sequences from Escherichia coli strain 0157:H7, useful for
PT      detecting E. coli 0157:H7 in a sample, or in designing diagnostic probes
PT      which can be used to distinguish strain 0157:H7 from strain K12 using
PT      molecular techniques.
XX      Claim 16; SEQ ID NO 13; 33pp; English.
XX      The invention relates to an isolated DNA molecule comprising an E. coli
CC      strain 0157:H7 sequence selected from a clostridial cytotoxin-like gene,
CC      a urease gene cluster, a RTX toxin-like gene cluster, a locus of
CC      enterocyte effacement and 2 genes from its associated lymphocytic phage
CC      933W (a putative serine/threonine kinase and a tail fibre gene). E. coli
CC      0157:H7 can cause food poisoning, specifically acute haemorrhagic colitis
CC      (which can develop into haemolytic uraemic syndrome). Also included are
```

```
CC      an isolated DNA molecule comprising a nucleotide sequence identical to at
CC      least 25 contiguous nucleotides contained in DNA sequences selected from
CC      ACD19988-ACD19242 (being 255 E. coli 0157 DNA sequences which are not
CC      found in E. coli K12), a recombinant DNA construction comprising the DNA
CC      above and a method for detecting E. coli 0157:H7 (ATCC 43895) in a sample
CC      (or distinguishing between 0157 and K12) using a probe derived from one
CC      of the 255 sequences. The DNA sequences are useful in detecting E. coli
CC      0157:H7 in a sample, for the early diagnosis of humans and livestock
CC      infected with 0157:H7, and in designing diagnostic probes which can be
CC      used to distinguish strain 0157:H7 from strain K12 using molecular
CC      techniques. The present sequence is one of the 255 E. coli 0157:H7 DNA
CC      sequences (termed OZ1D_1-OZ1D55). Note: The sequence data for this patent
CC      did not form part of the printed specification, but was obtained in
CC      electronic format directly from the USPTO at
CC      seddata.uspto.gov/sequence.html?docid=20030023075 (updated on 27-OCT-2003
CC      to standardise OS field)
SQ      Sequence 10523 BP; 2777 A; 2306 C; 2447 G; 2993 T; 0 U; 0 Other;

Query Match      3.2%; Score 75; DB 8; Length 10523;
Best Local Similarity 80.4%; Pred. No. 2.6e-08;
Matches 115; Conservative 0; Mismatches 20; Indels 8; Gaps 2;

QY      2168 CCGCTCCTCGCTGATACAGACATTCACGTATGAGGGAAGGTATGATGATGCTGGG 2227
Db      10258 CCAATCCTTCTCATATACAGGCAATTCCTTCACTGACATGGAAGAAAGTATGATCTGGA 10317
QY      2228 TGTGGGGAAGTCGTGAAGAAAGAAAGTCACTGCTGCTTGTGACATCATCTGATCTTC 2287
Db      10318 -----GAATTCGTGAAGAAAGAAAGAAAGATGCTGCTCGCGCTTTGTCGATGTTATCTTC 10371
QY      2288 TTACTGTTATGCAAGTGTAGT 2310
Db      10372 --ATTGGCTATGCAAGTCGTAA 10392

RESULT 6
ADCO1049 standard; DNA; 23480 BP.
XX      ADCO1049;
AC      ADCO1049;
XX      04-DEC-2003 (first entry)
DT      Enterohaemorrhagic E. coli 0157:H7-specific nucleic acid SEQ ID NO: 1093.
DE      ds; gene; enterohaemorrhagic; anti-bacterial.
XX      Escherichia coli; 0157:H7.
XX      JP2002355074-A.
XX      10-DEC-2002.
XX      24-JAN-2002; 2002JP-00015959.
XX      24-JAN-2001; 2001JP-00112010.
XX      (UYTS-) UNIV TSUKUBA.
XX      WPI; 2003-451640/43.
XX      Enterohaemorrhagic Escherichia coli 0157:H7-specific nucleic acid molecule
PT      and a polypeptide and its use, a polypeptide, a vector and a host cell.
PT      Claim 2; SEQ ID NO 1093; 2067pp; Japanese.
XX      The invention relates to a novel enterohaemorrhagic Escherichia coli
CC      0157:H7-specific nucleic acid molecule. A polynucleotide of the invention
CC      has anti-bacterial activity. The polypeptide can be used in detection
CC      and/or treatment of 0157:H7 infection. The nucleotide sequence of the
CC      genome of Enterohaemorrhagic E. coli 0157:H7 was determined. The present
CC      sequence represents an E. coli 0157:H7-specific nucleic acid of the
```


CC invention.

SQ Sequence 23480 BP; 6393 A; 4893 C; 5323 G; 6871 T; 0 U; 0 Other;

Query Match 3.2%; Score 75; DB 9; Length 23480;
Best Local Similarity 80.4%; Pred. No. 3.5e-08;
Matches 115; Conservative 0; Mismatches 20; Indels 8; Gaps 2;

OY 2168 CCGCTCCTCGGCTCAACAGACATTCACATACAGTTATGGCGGAAGGTATGCATGCTGGG 2227
|||
DB 10265 CCAATTCCTTGCTCATTAACAGGCAATTACTTCACTGATCAGTAGGAGAAAGTATGATGCTGGA 10324
|||||

OY 2228 TGTCGGGGAAGTCCTGGAAGAAGAAAGTCAAGCTGCCCTGTTTGACATCACTGCATCTTC 2287
-----SAAAGTCGTGAAGAAGAAAGAACATGCTGCGCGCTTTGTCTTCACGTTTATCTTC 10378
|||||

OY 2288 TTACTGGTTATGCAGTCTGACT 2310
|||||

Db 10379 --ATTGCTATGCAAGTCGTAA 10399

RESULT 7
ABS78855
ID ABS78855 standard; DNA; 9393 BP.
XX
AC ABS78855;
XX
DT 17-DEC-2002 (first entry)
XX
DE E. coli CFT073 genomic sequence #22.
XX
RW Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KM urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; atrophic; nephrotropic; gene; de.
XX
OS Escherichia coli.
XX
PN MO200259320-A2.
XX
PD 01-AUG-2002.
XX
PF 19-OCT-2001; 2001WO-US046833.
XX
PR 19-OCT-2000; 2000US-0242412P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Blatner FR, Welch RA, Burland VD;
XX
DR WPI; 2002-691532/74.
XX
PT New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
PT useful for preventing or treating E. coli CFT073 infection in humans or
PT livestock.
PS Claim 1; Page 105-110; 765pp; English.
XX

CC The present invention relates to polynucleotide sequences from the genome
CC of the pathogenic Escherichia coli strain CFT073. Almost all the
CC sequenced present in E. coli CFT073 are absent in the previously
CC sequenced laboratory strain K-12. The polynucleotide sequences of the
CC invention are useful for preventing, diagnosing or treating E. coli
CC CFT073 infection in humans or livestock. The polynucleotide sequences are
CC useful for preventing urinary tract infections and pyelonephritis.
CC Likewise, the polypeptides encoded by the different open reading frames
CC (ORF1-5) are useful for generating a vaccine against uropathogenic E.
CC coli strains. ABS78834-ABS79085 represent genomic sequences from E. coli
CC strain CFT073
XX

SQ Sequence 9393 BP; 2671 A; 2034 C; 2362 G; 2322 T; 0 U; 4 Other;
Query Match 3.1%; Score 71.4; DB 6; Length 9393;
Best Local Similarity 71.5%; Pred. No. 2.1e-07;

Matches	108;	Conservative	0;	Mismatches	41;	Indels	2;	Gaps	1;
QY	2167	TCGCGTCTCGCTATATTAACAGACATTCACTACATATATGCGGGAAAGATATCATGCTGG							2226
Db	8921	TCGCCAGTTCGCTCATATATAGACATTCACACTATATTATAGAAAGATATGCATGCTGCT							8980
QY	2227	GTGAGGGGAATCGTGAAGAAAGAAAGACTAGCTGCGCTGTTGACATCACTGCTATCTT							2286
Db	8981	GAGTGGGGAAAGTGTGAAAGAAAGAAAGACAGCGCTTGTGTGCTGCACGTTATCTT							9040
QY	2287	CTTACTGGTATTCAGGCTCGTACGAGGGGTGG							2317
Db	9041	C-ATTGGCTATGCAGAGTCTGTATATCAAGGC							9069

xx	RESULT 8	
xx	ADA71938/C	
ID	ADA71938 standard; DNA; 2000 BP.	
xx		
xx	ADA71938;	
xx		
DT	20-NOV-2003 (first entry)	
xx		
DE	Rice gene, SEQ ID 5263.	
xx		
xx	Plant; bacterial infection; fungal infection; viral infection; rice;	
xx	gene; ds.	
OS	Oryza sativa.	
xx		
xx	WO2003000898-A1.	
PD		
xx	03-JAN-2003.	
xx		
PF	22-JUN-2001; 2001WO-IB001105.	
xx		
PR	22-JUN-2001; 2001WO-IB001105.	
xx		
PA	(SYGN) SYNGENTA PARTICIPATIONS AG.	
xx		
PI	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;	
PI	Karagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;	
xx		
DR	WPI; 2003-175290/17.	
xx		
PT	Identifying at least one gene involved in plant resistance or response to	
PT	pathogenic infection for conferring resistance or tolerance to a plant to	
PT	bacterial, fungal or viral infection by determining or detecting plant	
PT	gene expression.	
xx		
xx	Claim 27; SEQ ID NO 5263; 899pp; English.	
xx		
CC	The present invention relates to a method (M1) for identifying genes	
CC	involved in plant resistance or response to pathogenic infection. M1	
CC	comprises identifying a gene whose expression is significantly altered in	
CC	the incompatible interaction of plant gene expression relative to	
CC	expression of the gene in an uninfected plant, in a mutant plant that	
CC	does not express a gene associated with response to pathogenic infection,	
CC	or in a corresponding incompatible or compatible interaction. (M1) is	
CC	useful for conferring resistance to resistance or tolerance to a plant to	
CC	bacterial, fungal or viral infection. The present sequence was used to	
CC	illustrate the invention.	
xx		
xx		
SO	Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;	
Query Match	2.2%; Score 50.8; DB 7; Length 2000;	
Best Local Similarity	12.0%; Pred. No. 0.02;	
Matches	43; Conservative 164; Mismatches 151; Indels 0; Gaps 0	
Oy	748 ATATTGTCATDAAACAATTAGTGAATTATGATCGTCGCACTGTGATCTGGTTCATTA 807	
Db	1014 ATYTRNSMSYTWYAAKKYTKWTATVASTWKWYAWKRAAYAMSRSKTYWCTGKEMATYC 955	

QY 1363 CCATTACAC 1375
|||
Db 151 AATTAAACAC 139
|||
RESULT 12
ID AAK78745 standard; DNA; 5739 BP.
AAK78745/c
AC AAK78745;
XX
XX 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33557.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
XX PN WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.

PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 11-OCT-2000; 2000US-0239935P.
PR 11-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.


```
DB          5713 ATATTTAACACG 5701

RESULT 14
AAK78746/C
ID AAK78746 standard; DNA; 33513 BP.
XX
AC AAK78746;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33558.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 11-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0217496P.
PR 26-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226641P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231249P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.

PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234979P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246529P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246533P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249257P.
PR 17-NOV-2000; 2000US-0249259P.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 6, 2004, 05:01:45 ; Search time 9045 Seconds

(without alignments) 11126.868 Million cell updates/sec

Title: US-10-074-744-4

Sequence: 1 aagctttcttaatttaacct.....gtcgtatggtgtgacacaca 2322

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_by:*

13: gb_un:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_of:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_sy:*

39: em_hugo_hum:*

40: em_hugo_mus:*

41: em_hugo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	232	100.0	48502	7 LAMCG	J02459 Bacteriophage
2	1369	59.0	7144	12 AF108217	AF108217 Cloning v
3	1369	59.0	7232	12 AF179893	AF179893 Cloning v
4	1260	54.3	133888	2 AC010769	AC010769 Homo sapi
5	947	40.8	42529	12 CVU3284	CVU3284 Cloning vec
6	947	40.8	42530	12 CVU3285	CVU3285 Cloning vec
7	947	40.8	42531	12 CVU3286	CVU3286 Cloning vec
8	947	40.8	42704	12 CVU37692	CVU37692 Cloning vec
9	486.2	20.9	2242	8 RSSTR1	X62334 R. serpentin
10	329.8	14.2	10124	1 AE006063	AE006063 Pasteurel
11	201.4	8.7	31067	1 AE016869	AE016869 Pseudomon
12	142.2	6.1	150050	1 MLO672112	AL672112 Mesorhizo
13	142.2	6.1	348580	1 BX572601	BX572601 Rhodosphe
14	104.8	4.5	2365	1 SLHAUR	X77574 S. lividans
15	75	3.2	10523	6 AR204117	AR204117 Sequence
16	75	3.2	12670	1 AE005203	AE005203 Escherich
17	75	3.2	23480	6 BD184783	BD184783 Nucleic a
18	75	3.2	295741	1 AP002551	AP002551 Escherich
19	71.4	3.1	9393	7 PH22XISA	PH22XISA Bacterioph
20	71.4	3.1	9393	6 AX702436	AX702436 Sequence
21	71.4	3.1	306358	1 AE016759	AE016759 Escherich
22	65	2.8	492	6 AR376208	AR376208 Sequence
23	53.6	2.3	11565	8 AP029891	AP029891 Cryphonc
24	52.8	2.3	302156	3 AC116977	AC116977 Dictyoste
25	52.2	2.2	128683	2 AC135793	AC135793 Oryza sat
26	51.6	2.2	1141	6 AX083744	AX083744 Sequence
27	50.8	2.2	2000	6 AX655393	AX655393 Sequence
28	49.8	2.1	2000	6 AX655393	AX655393 Sequence
29	49	2.1	183700	2 AC068612	AC068612 Homo sapi
30	49	2.1	187033	6 AC092965	AC092965 Homo sapi
31	48.4	2.1	1141	6 AX083744	AX083744 Sequence
32	48.2	2.1	303414	1 AE015938	AE015938 Clostridi
33	48	2.1	1302	9 AF090946	AF090946 Homo sapi
34	47.8	2.1	408	6 AX300942	AX300942 Sequence
35	47.4	2.0	349751	3 PFMA4P3	AL035476 Plasmodu
36	47.2	2.0	251762	3 AE014851	AE014851 Plasmodu
37	47.2	2.0	256172	2 AC005139	AC005139 Plasmodu
38	47.2	2.0	300050	1 AP004171	AP004171 Mycoplasma
39	47.2	2.0	310779	2 AC005140	AC005140 Plasmodu
40	46.8	2.0	721	3 AMU35790	U35790 Anopheles m
41	46.8	2.0	55247	9 AL136134	AL136134 Human DNA
42	46.8	2.0	115859	8 CNU08CE0	BX000510 Oryza sat
43	46.8	2.0	212685	5 BX294380	BX294380 Zebrafish
44	46.6	2.0	2661	3 AF030966	AF030966 Schistosom
45	46.4	2.0	140777	8 AP005783	AP005783 Oryza sat

ALIGNMENTS

RESULT 1

LOCUS LAMCG 48502 bp DNA circular PHG 17-APR-2002

DEFINITION Bacteriophage lambda, complete genome.

ACCESSION J02459 M17233 M24325 V00636 X00906

VERSION J02459.1 GI:215104

KEYWORDS DNA-binding protein; circular; coat protein; complete genome; origin of replication; repressor; unidentified reading frame.

SOURCE Bacteriophage lambda

ORGANISM Bacteriophage lambda

REFERENCE 1 (bases 1 to 12)

AUTHORS Wu, R. and Taylor, B.

TITLE Nucleotide sequence analysis of DNA. II. Complete nucleotide

[illegible]


```

/product="chloramphenicol acetyltransferase"
/protein_id="AA01200.1"
/db_xref="GI:6010637"
/translation="MEKKITGYTVDSIQSHRKHEKFAFQSVACQTYNQVQDLITAF
LKYVKKRKHRYPAFPHILRLKMAHFEFFAKDGLVIMDSVHPCTVPHRETEFT
SSIMSEYHDFRQPLHYSQDVACYGSLVYFPKGIENNEFVSANWFSFEDLVN
ANNDNFPAFPIFMKYVTOQDKVIMPLAIOVHNAVCDGFHVRMLNELQOYCDWQGG
A"
complement(6231..6285)
/gene="cat"
/note="pl promoter"
/evidence=experimental
/complement(6240..6256)
/gene="cat"
/bound_molety="O-L1 repressor"
/complement(6264..6280)
/gene="cat"
/bound_molety="O-L2 repressor"
6586..7119
/gene="gat"
6586..7119
/gene="gat"
/codon_start=1
/evidence=experimental
/translation="1"
/product="gentamicin acetyltransferase"
/protein_id="AA01201.1"
/db_xref="GI:6010638"
/translation="MLRSSNDVYQGGSPRTKUGSSMGIIIRCRLEPDDQVSKRNL
DLREREDVAITISQHPDSYDGNLRSKTFILALPDGDAVGAALAVLEFEDP
RSEIYLDLAVSGEHRQGIATLILNLRKHAANLGAIVYVQADYDGDPAVALYTL
GIREVWHPIDESTAT"

```

ORIGIN

Query Match 59.0%; Score 1369; DB 12; Length 7232;

Best Local Similarity 100.0%; Pred. No. 2.8e-276; Indels 0; Gaps 0;

Matches 1369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 954 TAAAGCGAGATCGCTAGTATTTTAAACTATGCTGGCAGCATTTCTGAATCAATATA 1013
DB 1 TAAAGCGAGATCGCTAGTATTTTAAACTATGCTGGCAGCATTTCTGAATCAATATA 60
QY 1014 AAAGTATGTGATCCTTTGCTGGGTCAGGTTGTTCTTTAGAGAGATAAAGATCAAA 1073
DB 61 AAAGTATGTGATCCTTTGCTGGGTCAGGTTGTTCTTTAGAGAGATAAAGATCAAA 120
QY 1074 TGCACCTAAAGCAAACTGAAACAGCATCGAAATATCCCTTGGGATTCCTTGACCTGAT 1133
DB 121 TGCACCTAAAGCAAACTGAAACAGCATCGAAATATCCCTTGGGATTCCTTGACCTGAT 180
QY 1134 AAGTCTATTTATTTCAAGAGAAAATATTCATGTTTCTGGGTTGGTGAATGACCAAT 1193
DB 181 AAGTCTATTTATTTCAAGAGAAAATATTCATGTTTCTGGGTTGGTGAATGACCAAT 240
QY 1194 CATTCATTCAAAATGTTGTTTACACACACCATTCGCGCCGATTAAGAGATGAATGTT 1253
DB 241 CATTCATTCAAAATGTTGTTTACACACACCATTCGCGCCGATTAAGAGATGAATGTT 300
QY 1254 CGTGTGGGCAATAGATTAACCGTCACCTCAAAAGTATAGTTAATCACTGAATCCGGG 1313
DB 301 CGTGTGGGCAATAGATTAACCGTCACCTCAAAAGTATAGTTAATCACTGAATCCGGG 360
QY 1314 AGCATTTTCTATTAATGAAGAGTGAATCTGACATTTCTGGCAAAACCATTTAAACAC 1373
DB 361 AGCATTTTCTATTAATGAAGAGTGAATCTGACATTTCTGGCAAAACCATTTAAACAC 420
QY 1374 ACGTGCAGACTGTCATGAATTTCTGAAAGATTACCCCTTAAGTAATGAGGTGTTAAG 1433
DB 421 ACGTGCAGACTGTCATGAATTTCTGAAAGATTACCCCTTAAGTAATGAGGTGTTAAG 480
QY 1434 GAGCGTTTCATTTCAATGTCGCTATTCATGCTTTGGCATCTACTAATTCCTGAATAGC 1493
DB 481 GAGCGTTTCATTTCAATGTCGCTATTCATGCTTTGGCATCTACTAATTCCTGAATAGC 540

```

```

QY 1494 TTTAAGAGGTTATGTTTAAACCATCGCTTAATTTGCTGAGATTAACATAGTACAT 1553
DB 541 TTTAAGAGGTTATGTTTAAACCATCGCTTAATTTGCTGAGATTAACATAGTACAT 600
QY 1554 GCTTTCACCTAAGAGAAAACCATTTCAAGGAGTTGACATGATTTTATCTATTAATGA 1613
DB 601 GCTTTCACCTAAGAGAAAACCATTTCAAGGAGTTGACATGATTTTATCTATTAATGA 660
QY 1614 ATTAAGGCTTACTCTCTTTTGAAGCTTAAACCAATTTTAAACATTTGCGATATGCA 1673
DB 661 ATTAAGGCTTACTCTCTTTTGAAGCTTAAACCAATTTTAAACATTTGCGATATGCA 720
QY 1674 TTTTTCACCATGCTCATCAAGACAGATAGATTAACATTTGTAACAAAGATATGATTT 1733
DB 721 TTTTTCACCATGCTCATCAAGACAGATAGATTAACATTTGTAACAAAGATATGATTT 780
QY 1734 CCAACCATCTGCTCGTGAAGATGCTTATTTTCTACAGCAAGATATACCGGCTCT 1793
DB 781 CCAACCATCTGCTCGTGAAGATGCTTATTTTCTACAGCAAGATATACCGGCTCT 840
QY 1794 TTCAATTAACAATAACCTCAACATATAGTAACCTTAATTTTATTAATTAACCGCAAT 1853
DB 841 TTCAATTAACAATAACCTCAACATATAGTAACCTTAATTTTATTAATTAACCGCAAT 900
QY 1854 TATTTGGCGGCAACACAGATCTCTCTTTTAAAGTTACTCTATATACATAGTTTCAT 1913
DB 901 TATTTGGCGGCAACACAGATCTCTCTTTTAAAGTTACTCTATATACATAGTTTCAT 960
QY 1914 CTAAATAATAGTAGATTTGAATTTAAACGGGACATCGATTTGATTTTCCATATTTAGCT 1973
DB 961 CTAAATAATAGTAGATTTGAATTTAAACGGGACATCGATTTGATTTTCCATATTTAGCT 1020
QY 1974 TTCTGCTTCTCTTTGATTAACCCCATCTGTTATTAATGATGTCATGTCATGTCATTTAACA 2033
DB 1021 TTCTGCTTCTCTTTGATTAACCCCATCTGTTATTAATGATGTCATGTCATGTCATTTAACA 1080
QY 2034 ACGATATAGTCTATTAATGATATATAGTATGCGGCAAGATTAAGCTTTCAGGCTTCTG 2093
DB 1081 ACGATATAGTCTATTAATGATATATAGTATGCGGCAAGATTAAGCTTTCAGGCTTCTG 1140
QY 2094 AAGAAGGCTTTCAAGTCTAATTAAGCCGATATAGTATGCGGCAAGCTTTCAGGCTTCTG 2153
DB 1141 AAGAAGGCTTTCAAGTCTAATTAAGCCGATATAGTATGCGGCAAGCTTTCAGGCTTCTG 1200
QY 2154 ATTAAGTCTAATCTTCCCTCTGCTCATTAACAGACATTTCACTACATTAATGCGGAAAG 2213
DB 1201 ATTAAGTCTAATCTTCCCTCTGCTCATTAACAGACATTTCACTACATTAATGCGGAAAG 1260
QY 2214 GATATGATGCTGCTGCTGCGGAAAGTCTGTAAGAGAAAGAGTACGCTGCTGTTGACA 2273
DB 1261 GATATGATGCTGCTGCTGCGGAAAGTCTGTAAGAGAAAGAGTACGCTGCTGTTGACA 1320
QY 2274 TCACTGCTATCTCTTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2322
DB 1321 TCACTGCTATCTCTTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369

```

RESULT 4

AC010769/c 133888 bp DNA linear HTG 07-MAY-2001

LOCUS AC010769 Homo sapiens chromosome 15 clone RP11-28B17 map 15, WORKING DRAFT

DEFINITION SEQUENCE, 11 unordered pieces.

ACCESSION AC010769.3 GI:7341761

VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 15, clone RP11-28B17


```

AGQLWLTVRVVDNATAMSEAGHISAMQWRLENI SVTLPAASHAI PHLTSEMDFC
IBENKRMQFNRQSGFSLQWIMIGDKQLLPDQFTRAPLDNDIGSEATRIDPNAM
VEBWKAGHQAQOAAALLOCTAPLADVLITTAHMOHQKTEISRKYRISGSGOM
AITVDEVAASDTPHAPRIGLNCQAOABVWMDLGGPOBNYDRLTAACTFDWDLPL
SDMYTPYVPSBENGSLACRGRELNGHPQWDFOPNRSISQOQLWETSIRHLHAE
GTMLNTIDGFRHMGIGSDMSVPSYSAEQLSAGRIHYQLWVCQR
misc_feature
  complement (23875..23958)
  /note="multiple cloning site"
  complement (23974..24142)
  /note="rimBT terminator"
  24476..25666
  /gene="tetr"
  /gene="tetr"
  24476..25666
  /gene="tetr"
  /codon_start=1
  /transl_table=11
  /product="tetracycline resistance protein"
  /protein_id="AAC53653.1"
  /db_xref="GI:1066311"
  /translation="MKSNNALIVILGTVTLDAVGIQVMEPLRLNDIVHSDSIASH
  YGVLALYALMOFLCAPVLGALSDRFGRFVLLASLIGATIDVAINATPVLITVAG
  RIVAGITGATGAVAGAYIADITGDEBARHFGMSACFGVAGVAPVAGGLGLAISLH
  APRLAAYVINGNLILIGCFILMOBSHGERPMPLEFNPVSPRMAPGMTVAILMTV
  PRIMQVGVGPALMWTREGDRFRMSMTIGSLAPGILHALAQAFVTPAPKRRGE
  KQALIGMAADALGVLLAFATRGMAFPMTIILASGIGMPALQAPLISQVDDHOG
  QLOGSLAALTSITIGPLIVTAIYASASTNGMLMTVGAALYLVCLPALRRGAMSR
  ATST"
misc_feature
  33603..35451
  /note="imm21 substitution"
  33603..35451
  /note="imm21 substitution"
ORIGIN
Query Match      40.8%; Score 947; DB 12; Length 42530;
Best Local Similarity 100.0%; Pred. No. 4.9e-188;
Matches 947; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 1376 GTGCGAAGCTGCTCATGAAATTTTGAAGAAGTTACCCCTCAAGTAAGAGGTGTAAAGA 1435
  DB 25764 GTGCGAAGCTGCTCATGAAATTTTGAAGAAGTTACCCCTCAAGTAAGAGGTGTAAAGA 25823
  QY 1436 CGCTTCATTTTCAATGCGGCTAATGATTTGGCCACTACTAATTCCTGAATAGCTT 1495
  DB 25824 CGCTTCATTTTCAATGCGGCTAATGATTTGGCCACTACTAATTCCTGAATAGCTT 25883
  QY 1496 TAAGAAGGTTATGTTTAAACCATGCTTAATTTGTGAGATTAACATAGTAGTCAATGC 1555
  DB 25884 TAAGAAGGTTATGTTTAAACCATGCTTAATTTGTGAGATTAACATAGTAGTCAATGC 25943
  QY 1556 TTTTCACCTAAGAAAAAATTTTCAGGAGTTGACTGAATTTTTCATTTAATGAAT 1615
  DB 25944 TTTTCACCTAAGAAAAAATTTTCAGGAGTTGACTGAATTTTTCATTTAATGAAT 26003
  QY 1616 AAGTGCTTACTCTCTCTTTTGAACCTAACAAACCAATTTTAAACATTTCCGATATCGACTT 1675
  DB 26004 AAGTGCTTACTCTCTCTTTTGAACCTAACAAACCAATTTTAAACATTTCCGATATCGACTT 26063
  QY 1676 TTTTCACCATGCTCATCAAGACAGTAAGATTAACATTTGTAACAAAGAAATGTCATTC 1735
  DB 26064 TTTTCACCATGCTCATCAAGACAGTAAGATTAACATTTGTAACAAAGAAATGTCATTC 26123
  QY 1736 AACCATCTGCTCGAGGAATGCTTATTTTCTACTGCGAGAAATATACCCGCTCTTTT 1795
  DB 26124 AACCATCTGCTCGAGGAATGCTTATTTTCTACTGCGAGAAATATACCCGCTCTTTT 26183
  QY 1796 CAATTAACACTTAACTCCAAACATATAGTAACCCCTAATTTTAAATTAACCGCAATTTA 1855
  DB 26184 CAATTAACACTTAACTCCAAACATATAGTAACCCCTAATTTTAAATTAACCGCAATTTA 26243
  QY 1856 TTTGGCGGCAACACAGAGATCTCTTTTAAAGTTACTCTTAATTAACATACGTTTTCATCT 1915
  DB 26244 TTTGGCGGCAACACAGAGATCTCTTTTAAAGTTACTCTTAATTAACATACGTTTTCATCT 26303
  QY 1916 AAAAATTAGTAGTATTTGAATTAACGGGGCATGCTATTTAGTTTTCATATTTAGCTTT 1975

```

```

DB 26304 AAAAATTAGTAGTATTTGAATTAACGGGGCATGCTATTTAGTTTTCATATTTAGCTTT 26363
QY 1976 CTGCTTCCTTTTGGATTAACCACTGTAATTCATGTCATGTCATGTCATGTTTATACCAAC 2035
DB 26364 CTGCTTCCTTTTGGATTAACCACTGTAATTCATGTCATGTCATGTCATGTTTATACCAAC 26423
QY 2036 GATATAGTCTATTAAATGATATATATAGTATGCGGACAGATTAAGCTCTTCAGGCTTCGAA 2095
DB 26424 GATATAGTCTATTAAATGATATATATAGTATGCGGACAGATTAAGCTCTTCAGGCTTCGAA 26483
QY 2096 GAAGCGTTTCAATTAATTAACCACTGTAATTCATGTCATGTCATGTCATGTTTATACCAAC 2155
DB 26484 GAAGCGTTTCAATTAATTAACCACTGTAATTCATGTCATGTCATGTCATGTTTATACCAAC 26543
QY 2156 AAGTGTAACTTCCGCTCTGCTCATTAACAGACATTCATTAAGTATGCGGAAAGT 2215
DB 26544 AAGTGTAACTTCCGCTCTGCTCATTAACAGACATTCATTAAGTATGCGGAAAGT 26603
QY 2216 ATGCATGCTGGGTGTGGGGAAGTCGTGAAGAAAGAAAGTCAGCTGCTGTTTGAATC 2275
DB 26604 ATGCATGCTGGGTGTGGGGAAGTCGTGAAGAAAGAAAGTCAGCTGCTGTTTGAATC 26663
QY 2276 ACTGCTATCTTCTTACTGCTTATGACAGGTGCTGTAAGTGGGTGGACACA 2322
DB 26664 ACTGCTATCTTCTTACTGCTTATGACAGGTGCTGTAAGTGGGTGGACACA 26710
RESULT 7
CUT93286 42531 bp DNA linear SYN 13-APR-1996
LOCUS CUT93286
DEFINITION Cloning vector TLP97-3, phage lambda lacZ translational fusion
ACCESSION U39286
VERSION U39286.1 GI:1066312
KEYWORDS Cloning vector TLP97-3
SOURCE Cloning vector TLP97-3
ORGANISM Cloning vector TLP97-3
REFERENCE 1 (bases 1 to 42531)
AUTHORS St Pierre, R. and Lim, T.
TITLE A refined vector system for the in vitro construction of
single-copy transcriptional or translational fusions to lacZ
JOURNAL Gene 169 (1), 65-68 (1996)
MEDLINE 96186904
PUBMED 8635751
REFERENCE 2 (bases 1 to 42531)
AUTHORS St Pierre, R.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-1995) Thomas Lim, Microbiology and Immunology,
University of Western Ontario, London, Ontario N6A5C1, Canada
FEATURES
source
  1..42531
  /organism="Cloning vector TLP97-3"
  /mol_type="genomic DNA"
  /db_xref="taxon:43842"
  /complement (19404..19489)
  /note="rimB2 terminator"
  /complement (19507..20760)
  /gene="lacY"
  /complement (19507..20760)
  /gene="lacY"
  /codon_start=1
  /transl_table=11
  /product="lactose permease"
  /protein_id="AAC53654.1"
  /db_xref="GI:1066313"
  /translation="MYLKYNTNFMBSLFFPPYFPIAGVAPPPIWILHNIHXSXD
  TGIIFPAISLFLFQPLFGLSDKLGRLRYLMTITGMVWAPPTFTFGLQYN
  IIVGSIYGVIFLISGCAALILAVLFFPAKTDAPSATVAVNAGHSAFSLALAEFR
  OPLKMFSLVIVGSCYDVYDQDFAPFTSFEATGCGTRVFGVYVTMGLNLSIM
  FPAFLINRIGRNALILACTIMSVRITGSSFATSAIEVLTIKTLHMEVPELIVGCF
  KYITSQFEVRSATIIYVCFQKQLAMTMSVLAGNMYESIGQGAVIVGLVALGF

```


FEATURES
source University of Western Ontario, London, Ontario 6N5A51, Canada
1. .42704
/location="Qualifiers"
/organism="Cloning vector lambda TXF97"
/mol_type="genomic DNA"
/db_xref="taxon:43300"
misc_feature
1. .42704
/note="EcoRI site normally present at this site has been removed by an uncharacterized mutation"
misc_feature
5505
/note="BamHI site normally present at this site has been removed by an uncharacterized mutation"
terminator
complement (19404..19489)
/note="trnBT2 terminator"
gene
complement (19507..20760)
/gene="lacz"
CDS
complement (19507..20760)
/gene="lacz"
/codon_start=1
/transl_table=1
/product="lactose permease"
/protein_id="AAC53645.1"
/db_xref="GI:1051182"
/translacion="MYTLKNTNFMWGLFFPPFPFMAGVPPPPPTIMHDINHSKD
TGIIFAIISLSPQLPGLSDKGLRKYLLMTITGMVMPAPFPIFGLPLYN
ILVGSIVGIVGFCFAGAPAVAFIEKYSRNSITFGMARMGVAGMALCASIVG
MPTINNOFVWLGSCALILAVLLFPAKTAPSSATYANVANGNSAFSLKLEFR
OPTLMFLSTVIGVSCVGVDOQFANFSPFATGCGVGVGTWMEELNMSAF
FPAFLINRIGCKNALLACTIMSVPFSSPATSALFYVITLTHMERPELLNGF
KYTTSQEVFPAATYVCCFCFQOLAMTMSVLAGMWSISGFQAYIVLGLVAGF
TLISVFTLSGPGSLRLQVNEVA"
complement (20812..23886)
/gene="lacz"
CDS
complement (20812..23886)
/gene="lacz"
/note="EcoRI site normally present at the end of lacz has been removed by an uncharacterized mutation"
/codon_start=1
/transl_table=1
/product="beta-galactosidase"
/protein_id="AAC53646.1"
/db_xref="GI:1051183"
/translacion="MTWITDLSAVLQGRDMPNGVTQNLRLAHPPEFASWNSSEAR
TDRPSQQLRSINGRFPAPBAPBESWLECDLPADTVVVPNSNMQMGYAPIT
NMYPIVTPNPPTENPTGCSLTENVDSEMLQEGTRIIFGVNSAFPLMCGRNV
GYGQDSRLPSEFPLSAFIRAGENRIAVLRMSGSLYLBNDMMRSGIFRDSILHK
PTQIISPHVATRENDPSRAVLBAVOMCGELRDYIRVTVSLMOGTOVASGTAAPG
GEIIDERGGYADRYTLRLNENPKLMSAEIPNIRAVVELHTADGTLIEBACDVGR
EVRILENGLLNGEPILIRGVNHEHHPHIGQVDEBTVQDILMLKQNFANVROCSH
YRHPILMYTLCDRGLYVDEBANIETHGVNPNRLNDPRLPAMSERVTRMVARDBN
HPSVITLSLNGESGHANDALYRWIKSVPSRPVQEGGAGATTADILICPMYARD
EDPVPFAPKMSIKMLSLFGETRPLILCYAAMSGSLGFAKYMOPFQYRLQOG
FMDPWVDSILIKYDENGPNWSAYGDEGDPNDROPOMGLVAFRPHPLBEAKOG
OCRFORPLSGCTLETVSEVLFRSDEMLHMMVALDCKPLASEVPLDVAPOCKOLIE
LPELPDESGAGLMTLVTVQPNRTAKSEGHISAKOFLAERLSTTLPAASHALPH
LTTSEMDFCILGNKRWQFNROSGPLSQMWIGDKQLTLPLRQFTPARHAGLISE
ATRIDPVAVERKKAAGHYQABALLOCTDITLADAVLITTAHMOGCTLPSRKT
YRIDSGOMATVDEVAADTPPARIGLNCOLAVERVMVGLGOEYVPRILTAA
CFPRMDPLSDMTVPYVPSENGRLGCTRELANTGPHOMRGDPFQNSISQOOLMERS
HRHLHAEBSTWNIQFHMGIIGDDSDMSVSAEPQLSAGRHYQVWQCK"
complement (24049..24132)
/note="multiple cloning site"
terminator
complement (24148..24316)
/note="trnBT1 terminator"
gene
24650..25840
/gene="tetr"
24650..25840
/gene="tetr"
CDS
24650..25840
/note="Description: tetracycline resistance gene;
tetracycline resistance protein"
/codon_start=1
/transl_table=1
/protein_id="AAC53647.1"

/db_xref="GI:1051184"
/translacion="MKSNNALVILGVTLDVAGLGLWPEVPLGLRDIVHSDSLASH
YGVLLALYALMPELCAPLGALSDRFRGRVLLASLIGATIDVAINATPVLIVLAG
RIVATGATGAVAGAYIADITDGEDRGRFMSACFQGVWAGVAGLGLAISIH
APFIAAYANGANLILGCEPLMOSHGERPMPRLPNPVSFRMRGMVITVALMTV
FRTMQLVGVPAALMVTIFGEDRFRMSATMIGSLAYFGIILHALQAFVTPATKRRE
KQAIAGMADADAGVYLAFAFGWMAFPMTIILASGGIGMPLQAMLSQVDDHOG
OLOGSLAALTSLTSYIGPLIVAIYAASASTWNLAMIVGALYVCLPALRGRGMSR
ATST"
misc_feature
31747
/note="EcoRI site normally present at this site has been removed by an uncharacterized mutation"
misc_feature
33777..35625
/note="imm21 substitution"
misc_feature
39168
/note="EcoRI site normally present at this site has been removed by an uncharacterized mutation"
ORIGIN
Query Match 40.8%; Score 947; DB 12; Length 42704;
Best Local Similarity 100.0%; Pred. No. 4.9e-188;
Matches 947; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1376 GTGCGAAGCTGCATGATATTTTGGAAAGACTTACCCCTTAAGTAATGAGGTGTTAAGA 1435
DB 25938 GTGCGAAGCTGCATGATATTTTGGAAAGACTTACCCCTTAAGTAATGAGGTGTTAAGA 25997
QY 1436 CGCTTCATTTTCATGATGCGGCTAATGATTTGGCCATACTAATCCGTAAGACTT 1495
DB 25998 CGCTTCATTTTCATGATGCGGCTAATGATTTGGCCATACTAATCCGTAAGACTT 26057
QY 1496 TAAAGAGTTATGTTTAAACCATCGCTTAATTTCTGAGATTACATGATGCAATGC 1555
DB 26058 TAAAGAGTTATGTTTAAACCATCGCTTAATTTCTGAGATTACATGATGCAATGC 26117
QY 1556 TTTCACCTAGGAAAAAATTTTTCAGGAGTTCAGTGAATTTTATCATTAATGAAT 1615
DB 26118 TTTCACCTAGGAAAAAATTTTTCAGGAGTTCAGTGAATTTTATCATTAATGAAT 26177
QY 1616 AAGTGCTTACTCTTCTTTTGAACCTAACAAACCAATTTTAACTTCCGATATGCAAT 1675
DB 26178 AAGTGCTTACTCTTCTTTTGAACCTAACAAACCAATTTTAACTTCCGATATGCAAT 26237
QY 1676 TTTCACCATGCTCATCAAGACAGTAAGATTAACATTTGTAACAAGAAATAGTCAATCC 1735
DB 26238 TTTCACCATGCTCATCAAGACAGTAAGATTAACATTTGTAACAAGAAATAGTCAATCC 26297
QY 1736 AACCATCTGCTGTAAGAAAGCCCTTATTTTCTATGAGAGAAATATACCCGCTCTT 1795
DB 26298 AACCATCTGCTGTAAGAAAGCCCTTATTTTCTATGAGAGAAATATACCCGCTCTT 26357
QY 1796 CAATTAACACTAACTCCAAATATAGTAACCTTAATTTTAAATAACCGCAATTTA 1855
DB 26358 CAATTAACACTAACTCCAAATATAGTAACCTTAATTTTAAATAACCGCAATTTA 26417
QY 1856 TTGGGCGGCAACACAGATCTCTTTTAAAGTTACTCTATTAACATAGCTTTTCAATCT 1915
DB 26418 TTGGGCGGCAACACAGATCTCTTTTAAAGTTACTCTATTAACATAGCTTTTCAATCT 26477
QY 1916 AAAAAATTAGTAGTATGAATCTTAACGGGGATGATAGTAGTTTGCATATTTAGCTTT 1975
DB 26478 AAAAAATTAGTAGTATGAATCTTAACGGGGATGATAGTAGTTTGCATATTTAGCTTT 26537
QY 1976 CTGCTTCCTTTGGATTAACCACTGTTATTCATGTTGCATGTTGACCTTTATACCAAC 2035
DB 26538 CTGCTTCCTTTGGATTAACCACTGTTATTCATGTTGCATGTTGACCTTTATACCAAC 26597
QY 2036 GATATAGCTTATTAATGATATATATATGATGCGGCAAGATTAAGCTCTTCAAGCTTTGAA 2095
DB 26598 GATATAGCTTATTAATGATATATATATGATGCGGCAAGATTAAGCTCTTCAAGCTTTGAA 26657
QY 2096 GAAGCGTTTGAAGTAAATTAAGCGGATAGTATGCGACGGAAGCTTGGATTTTTCAT 2155

Db 2658 GAAGGTTTCAAGTACTAATAGCCGATAGATAGCCAGGACTTGTAGCCATTTTTCAT 26717

QY 2156 AAGTTAACTTCCGCTCTCTCCCTCAATAACAGACATTCATCAAGTTATGCGGAAAGGT 2215

Db 26718 AAGTTAACTTCCGCTCTCTCCCTCAATAACAGACATTCATCAAGTTATGCGGAAAGGT 26777

QY 2216 AAGCATGCTGGGTGGGGAAGTCCGTAAGAAAGAAATCGACGTCGCTTTTGACATC 2275

Db 26778 AAGCATGCTGGGTGGGGAAGTCCGTAAGAAAGAAAGAAATCGACGTCGCTTTTGACATC 26837

QY 2276 ACTGCTATCTTCTTACTGTTATGACAGTCTAGTGGTGACACACA 2322

Db 26838 ACTGCTATCTTCTTACTGTTATGACAGTCTAGTGGTGACACACA 26884

RESULT 9

RSSTR1 2242 bp DNA linear PLN 22-JUN-1992

LOCUS R.serpentina str1 gene for strictosidine synthase.

ACCESSION X62334.1 GI:21126

VERSION X62334.1

KEYWORDS str1 gene; strictosidine synthase.

SOURCE Rauwolfia serpentina (serpentwood)

ORGANISM Rauwolfia serpentina

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Gentianales; Apocynaceae; Rauwolfioideae; Vinceae; Rauwolfia.

1 (bases 1 to 2242)

Bracher,D. and Kutchan,T.M.

Strictosidine synthase from Rauwolfia serpentina: analysis of a gene involved in indole alkaloid biosynthesis

Arch. Biochem. Biophys. 294 (2), 717-723 (1992)

JOURNAL MEDLINE 92231602

PubMed 1567228

2 (bases 1 to 2242)

Kutchan,T.M.

Direct Submission

Submitted (27-NOV-1991) T.M. Kutchan, Lehrstuhl fuer Pharmazeutische Biologie, Universitaet Muenchen, Karlstr 29, D-W-8000 Muenchen 2

Location/Qualifiers

1..2242

/organism="Rauwolfia serpentina"

/mol_type="genomic DNA"

/db_xref="taxon:4060"

/cell_line="cell suspension culture"

/clone_lib="lambda charon 4A"

320..330

CAAT_signal

434..440

TATA_signal

466..4581

gene

/gene="str1"

466..>1581

prim_transcript

/evidence="experimental"

547..1581

CDS

/gene="str1"

/codon_start=1

/product="strictosidine synthase"

/protein_id="CAA44208.1"

/db_xref="GI:21127"

/db_xref="GOA:P15324"

/db_xref="SWISS-PROT:P15324"

/translation="MAKLSDSQMALFTVFLFLSSSLALSSPIIKELIIBAPYAPN SFTDSTNKGYTSVODGRVVKYEGNSGFVDFVAVASPYMKAFCEKSTDAEKPLCG RTYDISYLNQNNOLYIVDCYHLSVSGEGHATGLATSDVGVPKMYAVTVQRTG IYVFDVSTLDVDRGVQIIMPTSDKTRLIYDPSRTKETILLIKELIHYAGAAVSADS SFVIAEFLSHQIVKYMLEGKKAIVLYIKIPNGINKRNADGHPVSSSEBIDGMN HGRVDPKIKRDERENILFVILPPFAGEHFEQIQEDHGLNYIGTLPHSGVGLIYVD KKGNFVSSH"

polYA_signal

1672..1676

polyA_signal

1691..1695

misc_signal

1797..1804

ORIGIN /note="GT box"

Query Match 20.9%; Score 486.2; DB 8; Length 2242;

Best Local Similarity 97.3%; Pred. No. 1,8e-91;

Matches 505; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 927 ACAACCAATTAATGCTGAATGAATGATTTGAAGGAGATCGCCTAGTATTTAACTATT 986

Db 1703 AAAAAAAAAAAAAAAAAAGGAATTTAAGGGAGATCGCCTAGTATTTAACTATT 1762

QY 987 GCTGCACATTTCTTGAATTTCAATTAAGATTTGTACTTTTGTGGGTCAAGTTG 1046

Db 1763 GCTGCACATTTCTTGAATTTCAATTAAGATTTGTACTTTTGTGGGTCAAGTTG 1822

QY 1047 TTCTTTAGAGAGATTAAGATCAATGCACTAATCAAGAACTGAACCAAGATCGAAA 1106

Db 1823 TTCTTTAGAGAGATTAAGATCAATGCACTAATCAAGAACTGAACCAAGATCGAAA 1882

QY 1107 ATATCCCTTTGGGATTTCTTGAATTTCAATGATTTATTTTCAAGAAAAATATTCAT 1166

Db 1883 ATATCCCTTTTGGATTTCTTGAATTTCAATGATTTATTTTCAAGAAAAATATTCAT 1941

QY 1167 GTTTTCTGGTTGGTGGATTTGACCAATTCATTTCAATTTGTTTTCACACACC 1226

Db 1942 GTTTTCTGGTTGGTGGATTTGACCAATTCATTTCAATTTGTTTTCACACACC 2001

QY 1227 ATTCGCCCGATTAAGATGATGATGCTGCTGGGCAATGAATTAACCGTACCTCAAA 1286

Db 2002 ATTCGCCCGATTAAGATGATGATGCTGCTGGGCAATGAATTAACCGTACCTCAAA 2061

QY 1287 AGGTAACTTAATCACTGAATTCGGGAGCACTTTTCTATTAAATGAAGATGAATTC 1346

Db 2062 AGGTAACTTAATCACTGAATTCGGGAGCACTTTTCTATTAAATGAAGATGAATTC 2121

QY 1347 TGACAATCTGCGCAACCATTTAACAACGCGGCACTGCTCATTAATTTTCTGAAGAT 1406

Db 2122 TGACAATCTGCGCAACCATTTAACAACGCGGCACTGCTCATTAATTTTCTGAAGAT 2181

QY 1407 TACCCCTTAAGTATGAGTGTGTAAGACGCTTTCATT 1445

Db 2182 TACCCCTTAAGTATGAGTGTGTAAGACGCTTTCATT 2220

RESULT 10

AE006063 10124 bp DNA linear BCT 08-MAR-2001

LOCUS AE006063/c

DEFINITION Pasteurella multocida PM70 section 30 of 204 of the complete genome.

ACCESSION AE006063 AE004439

VERSION AE006063.1 GI:12720510

KEYWORDS SOURCE

ORGANISM Pasteurella multocida

Pasteurella multocida

Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales; Pasteurellaceae; Pasteurella.

1 (bases 1 to 10124)

May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whitlam,T.S. and Kapur,V.

Complete genomic sequence of Pasteurella multocida, PM70 Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)

JOURNAL MEDLINE 11248100

PubMed 1145866

2 (bases 1 to 10124)

Zhang,Q. and Kapur,V.

Direct Submission

Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA

Location/Qualifiers

1..10124

/organism="Pasteurella multocida"

/mol_type="genomic DNA"

/strain="PM70"
/db_xref="taxon:747"
78..458
/gene="PM0279"
78..458
/gene="PM0279"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02363.1"
/db_xref="GI:12720511"
/translation="MSNPVSWFEIHWNIQRATFYQAVFQLPITGDCPSGKMSVC
FSTIDAQYIGIMLYEDKDFPVVRGNNTLTFACEDCTVEARAVKFGSKLLQGKTS
IGDEGFVAIVEDSEGNRIIGHSMQ"
563..1729
/gene="sucC"
/note="synonym: PM0280"
563..1729
/gene="sucC"
/codon_start=1
/transl_table=11
/product="sucC"
/protein_id="AAK02364.1"
/db_xref="GI:12720512"
/translation="WMNLHEYQAKQIFAEYQLPVGKGYACKSADPADAIKRLNGDVAV
AKQVHAGRGKAGVGLVVRNEAEVRAFQDMQGLVTPQIDVNGQPVNTLYVEBS
SIRAEYLGAIVLRASORVVMSTEGEVNIIEVAEKTPHLHKMALIDPLTGMYPQG
RELAFKLGKGDQIKOPAHIFVOMAKFVEKDLLEVNPLVTKGNLILCLAKIYV
DSNALIKROPALKMOPDSQEDPREALAESQILVYALFENIGCMVNGAGLNMGTII
KLHGQGANPLDVGGAATKERVAAEFKILISDSVKAVALVNTFGIYRCDLIABGIIA
AVEVGVNVPVAVLRGNNAPIRGREILANSVNI IANTLTDAIQAIVKAECK"
1731..2600
/gene="sucD"
/note="synonym: PM0281"
1731..2600
/gene="sucD"
/codon_start=1
/transl_table=11
/product="sucD"
/protein_id="AAK02365.1"
/db_xref="GI:12720513"
/translation="WSILINKDKYVTCQGTGCGQTRHSQALAYGTQLVGVSPKCG
GTHHLGPVNTVADATGATASVTVYAPGCKALILEAIDAGIQLVICITEGTYT
LDMLQVKORLNTGVNMI GNCPCVITPEBCKIGIMGNHKKGIIGVRSGLTYE
AVKQTDDEGQSTCVGIGDPIGSGNFIIDILKLFODDPQTEAIVMIGEIGSAAEBA
AAVTKQVTKVYVSYAGVTAAPGKRMGHGALITISGKGTADKFALEAAGKTVAS
LARIQAATKALK"
2845..4439
/gene="PM0282"
2845..4439
/gene="PM0282"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02366.1"
/db_xref="GI:12720514"
/translation="MRFRLINRDSMPQNAQNVCLIDNMNDYSYITMFMSFEDSD
GVHEIGNIKIGFGQVETLPTAEIRNFRDGRVFDLPNFIISQGVDFYSRIEPL
SNVEKDIILKNLDVYVNPDLDBPQNEPVPYSIMHASEPITKQOFRLVKGKSP
TDTFSTIRNRQDSMGKITLEFEYKANSVSTINHAIGNGVGTITLAKMIAKSFN
KNYKDGAFYKKERIHFTNTPIKISDEYSRVAVSFVDFPIPNNDERKNTSYGL
KKSEDTQKEPKYFEEDFWLFLCEKRSFPRDLWVAIQNLSDNNPAMNLSHMLN
ECSKDILOFINRMSGHAFVLLITITQVAKVEKTLVILDEPESHLPPLSAFIR
TLSELNDRNAIIVATHSPVLOEIPKSCIMKIVRTGIEFKAVRLVETFGNIGIL
TRVFGLEVEKSGPHKLINESVNGCGYTESILSYNQQLKEGKLILKTLIRLDEKN
QGIFYDKD"
4358..5290
/gene="PM0283"
4358..5290
/gene="PM0283"
/codon_start=1
/transl_table=11
/product="unknown"

/protein_id="AAK02367.1"
/db_xref="GI:12720515"
/translation="MKKIKGFPIIMIKILPTDCQNSNIIIDSCNGILRDLRELTKII
VNGGILRLISKYIYAKUNNCLFERYKHICHSEIILMRGFPDSDLITCTMKLYEY
FTRKCGSYIDKILINAKDPLIOCPFCGGICVPSDELHPYKHYFAITSPFNVYVI
CKDCNKSCKNEFPYIEINKQILHPYLDQDCFFSBQMFNAVCIENNIASGTIVYVSF
PTNMSBDKKGKVFHEQFQFLDERFSIRAPYVLDIQAOTTRAKSKSLSEADCIETII
IDPIINSEYNNKHMVAVLYOAVKKNFTHIWNI"
complement (5372..6016)
/gene="adk"
/note="synonym: PM0284"
complement (5372..6016)
/gene="adk"
/codon_start=1
/transl_table=11
/product="adk"
/protein_id="AAK02368.1"
/db_xref="GI:12720516"
/translation="MKIILGAPGAGKTQAOPIKMKFGIPIQISTGDMRGAIKAGTD
LGRQAKTMDAGQVDPDLIISLVKSRVADQACKGFLDGFERTIPQADALATVGIQ
IDVLEFEDVEDVIVRMSSGRVHOASGRTHVAVNPKVEGDDVYGEDLIIRADDK
PETVLDRKLVHSTTKPLVDYVOAKAGNTKXFRLDGTKKVEVSQELDTIIA"
complement (6165..7442)
/gene="PM0285"
complement (6165..7442)
/gene="PM0285"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02369.1"
/db_xref="GI:12720517"
/translation="METQKPSLISQIFSRNMLCITFGSSGLPYLVLFQMPDAWSS
SNVDVKTIGFFFLIGFPYTWKFIISPLDRYENPFLGRSRMMENTQVALALIALAG
QYDPVANITIVAVISTEIAVFSAYODIVIDAVREILITDELGNSIHNAVRISEL
IRGSLIILADSPLMDVFEIFETALFMIPLILSIPSKBEDVQIDSLPEYHFAIP
KREPSRKGYASALGLIFRFLYKRGDSATSIOTKRIIDMGSGKTDIGTVVILNSIM
ASTISMLGSLMLRIGINSLMLFGVQULITTLGFWMASYGFNNTIGSPERFVLT
VIIIEYIGVGLGAAFAVFAFMARETNPLTYTQALIFSLSAISPKIFSAFSGVIVADY
GYTFPFWICFFFAIPGMLMFKVAPMHSABORA"
complement (7445..8461)
/gene="galB"
/note="synonym: PM0286"
complement (7445..8461)
/gene="galB"
/codon_start=1
/transl_table=11
/product="galB"
/protein_id="AAK02370.1"
/db_xref="GI:12720518"
/translation="MALIVTGAAGYIGSHTVBELNANKVYVULDNCSNPSKLERV
AOITGRQVREYQGDILDTALLOKI FAEHQIQSVIHFAGLAVGSSVQKPAEYNNAT
GSLVLIQEMKAGVWNVFESSATVGDPIIPTBESCKVGTNPGTSGFMEVQIL
KOJAKAPERSITILREYENPVGAHESGLIGEDPNGIIPNNILPYISQVAILPLQLSYF
GSDYETHDGCQVADYIHVDLAIYGLKALDRHEDDAGLHLYNGTSGSYVLDVWVAF
EKYNDIKIIPYKLVDRRBDIATCYSDPSLAKTELMTAAAGLEGMMDTMMWCKKNK
GYRD"
8728..10041
/gene="PM0287"
8728..10041
/gene="PM0287"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02371.1"
/db_xref="GI:12720519"
/translation="MSLEKHPMTKRGSNVROEYIAGLTTFILAMVYSVIVPNNLGA
GPRABSVFATCIVAGIGSLILGLMANVPALICAIITAPFASPILVGOOVSIPVAL
GAVFLMGVVTIISATGIRAMILRNPSSTAHGAGIGTGLFILLIANGVGLVYNSQA
GLPVKMGELTSPFLVMSLIGLALILGLEKRVKGAIIIMVLIATLVGVDPNVKFG
SIFKMPFGENSFLLELDLGAOPALIPVFPALVMTAVDATGTRAVAGQANLLK
EGQIINGKALTDSSVSIIFSGLGTAPAVYIESAAGTAAGGTGLTAAVVGILFLI
MFPQPLPIVPGVATAPALMAYVGLMISVSKLMDPDPTIGAMSGVCAVPIVLTANI
VTGIMLGFALVIGRVVSGEYKKLNVGTVLIILALVAFYAFGNAL"

TITLE Nelson, W., Davidson, T., White, O., Fraser, C. and Colimer, A.
 JOURNAL Direct Submission
 Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
 Source Location/Qualifiers
 1. .310967

/organism="Pseudomonas syringae pv. tomato str. DC3000"
 /mol_type="genomic DNA"
 /strain="DC3000"
 /db_xref="taxon:223283"
 /note="pathovar: tomato"
 complement (283. .1092)
 /locus_tag="PSPPT03607"
 complement (283. .1092)
 /locus_tag="PSPPT03607"
 /note="The annotation of this multicopy IS21 family
 element is based in part on an HMM match, the ISfinder
 database, and on positional information.; similar to
 GP:6009454, and GP:17865161; identified by sequence
 similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="ISpey4, transposition helper protein"
 /protein_id="AA057081.1"
 /db_xref="GI:28854016"

/translation="MNLQHARLTCLCKLTERGVDMPHLAQQAAGSDSPADFLER
 LLAETPARSERSRQALKTALPAVLTQYDFATVGPRAQLCEIALSFVGRAB
 NITVGLSGVGKSLAIALAVRAVMAGITKTFVTAALMOLTAHROEALKEFSYV
 VMAKGLIDVIGLTPRGDEANLFFVAVAKRYEGSLITLSNLPFQMGATPADDT
 LTAAMDRLHHAHIVMTGESYRLKDKRAKGSASAPARKTEPBGON"

complement (1089. .2111)
 /locus_tag="PSPPT03608"
 complement (1089. .2111)
 /locus_tag="PSPPT03608"
 /note="The annotation of this multicopy IS21 family
 element is based in part on an HMM match, the ISfinder
 database, and on positional information.; similar to
 GP:14594862, and GP:17865160; identified by sequence
 similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="ISpey4, transposase"
 /protein_id="AA057082.1"
 /db_xref="GI:28854017"

/translation="MLTQEOSVEIKLAPRQGHGIFARELGISRNTVRYKARSL
 PSDKVPARPCKIDPFKDYLERLEKAPRPHIPTVLREITLALGSGSRLKAYLR
 PKRKAEPVPRFETLPGKOIQNDFTIRRGROIKAFVATTIGSRASFRSRRNS
 EAMITGKREAFYRGGVPEQALPDMAGTITITERAFEGGGRHPRILAADEGFLP
 KVCPYRAQTKGKVERFNGYKGSFTPLATLKSAGLTLDVVTANAHITQMDDEVAN
 QRHGTTGVQPAVALAOEQVLLPLPTQSLRPOAGLRGRVLPYESTLQHPISVYEQ
 LLEVRA"

2191. .3213
 /locus_tag="PSPPT03609"
 2191. .3213
 /locus_tag="PSPPT03609"
 /note="Warning: Expression or function of this gene might
 be disrupted by the first 2 residues of the coding sequence
 are contributed by the end-sequence of an insertion
 sequence, ISpey4.; identified by match to PFAM protein
 family HMM PF00589"
 /codon_start=1
 /transl_table=11
 /product="site-specific recombinase, phage integrase
 family"
 /protein_id="AA057083.1"
 /db_xref="GI:28854018"

/translation="MTTREGISRRSWESTGRALYFSPLOAHSLHMKVDKGEKD
 LVYGRDYCRDTCSLATSLRQRYVCEYDFSLKMKTEKLPYNESRVSRSFN
 SSRHPLASKTVLVQDVLPVLTKSLPRLRMSVEVELLTHAQNHHRMIRFALTTGR
 QDEIASPFLAVFENPEKGNKTSNIIQRLPDRDGTGKTRSKRDIIFISRFLELH
 AYLRMGESTLANSNAKELFLNRGEPYADFEKSI CRTLRNIGKKSIVSYSHMR
 HTVATQTLISLQKNSLIPLVFLQRLGHSSIQTTWYVLIHVNALADEAVIAYDELA
 NLSVA"

gene

CDS

3210. .4889
 /locus_tag="PSPPT03610"
 3210. .4889
 /locus_tag="PSPPT03610"
 /note="Identified by match to PFAM protein family HMM
 PF02899"

/codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA057084.1"
 /db_xref="GI:28854019"
 /translation="MTGKKTFTTDSIPAVNHNSTENQTKLTDPRPEYRVNDI
 GKATRRMTIDFERWTRADIDETIACQRIETVARQDDIDEPSTVASCKGGLRIF
 LDVYMLRAKAYEREELRADYRSLIDGFLGYNLDAI SRQSQSNFPAASILMALCA
 RGLIAYVSDSDHATFPKPNPTKTDGSPFKRQROEVAALKTAAPITKDDVVI
 TSSLSVALLTIALHGRNTTPLLMDRNLRAHPKONLPELVLMKRGHSTYSKTLR
 SPPDORALESTIPVATNTIROLIRVYHLTSGILSDAPEDIANRWITRGQAASARK
 VHELNDAMLKFAIQKYERDIDDTGKPLQLVNSRLKRPARKIRIYLLGNTDQET
 ALGNTPAVTEQDYLRLPTYESKRWKMGELVRLMTKITIGATFHPFVGGKRDPEEG
 QFAPKSGGVCFSLNCVRCKHYAVTADDLKLFSEFYRIYSRSSIGKTRWEKELSH
 IPRLDNYITIEELRRKIFKPEVEIARTARNTPHFMTCDALNSIEAEI"

gene

CDS

5227. .5260
 /locus_tag="PSPPT03611"
 /note="This region contains an authentic frame shift and
 is not the result of a sequencing artifact; This gene is
 interrupted by an IS element.; IS52, transposase,
 interruption-N"

5343. .5702
 /locus_tag="PSPPT03612"
 5343. .5702
 /locus_tag="PSPPT03612"
 /codon_start=1
 /transl_table=11
 /product="ISpey5, Orf1"
 /protein_id="AA057085.1"
 /db_xref="GI:28854020"
 /translation="MRRPDKVKRYVLYPKVPFRRKSIDGLAALVELIDIVAYVDPVL
 FLVFNKSRNRYKLLIWRNNGFCMLRLESERKISDDVDDEAVILTVQSLNMLDGF
 DLMRNPRIKILITRFVA"

gene

CDS

5720. .7273
 /locus_tag="PSPPT03613"
 5720. .7273
 /locus_tag="PSPPT03613"
 /note="Similar to GP:8037785; identified by sequence
 similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="ISpey5, transposase"
 /protein_id="AA057086.1"
 /db_xref="GI:28854021"

/translation="MRPLPNDLPDDPYLTKQMLSEYARSREGEVADAYATLVDLK
 EQIKLDRIFGRKSEQVEPNTPOLAFNEPESBPLPYGDADDEYVADAAPRRGKK
 PLASDLPRIEIHLEPHELTCAGGKGVSETSQDLIVPQIIVIKHIRVYVG
 RGEASAPVTAGKPAOLIEKMSAPSVIAMLITTKYVNGLLPRLRETYLSHNGEIPPO
 TLRWIIQCGEHIQPLINLRDRLBESPVYHCBETRYOYKPEPRDTSQSMVQVS
 GPPDHKVLFDYTSRAQDVPLRLBESRYVWTDVYAGNALALQGEVRLAMAA
 RRRFVEAQKQPKPKGRADIALTMNKYIGIERLDGSDQEFTRQESLPIILQ
 LKSMLDQSHVPOSYLGKAVGYLANMGRILKRYVAGFLPIDNNAERAIKPFVIG
 RKAMLFEDTQKATASAOIYSLVETAKVNGEPYTWLRVHLERPLAOSVADYBALLP
 WNCSPEMBR"

gene

CDS

7321. .8271
 /locus_tag="PSPPT03614"
 /note="This region contains an authentic frame shift and
 is not the result of a sequencing artifact; This gene is
 disrupted by an IS element.; IS52, transposase,
 interruption-C; similar to GP:8885851, and GP:1196750;
 identified by sequence similarity: putative"
 8452. .10053
 /locus_tag="PSPPT03615"
 8452. .10053
 /locus_tag="PSPPT03615"

gene

CDS

/note="Identified by match to PFAM protein family HMM
 PF00589"

REFERENCE	1 (bases 1 to 150050)	1 (bases 1 to 150050)
AUTHORS	Sullivan,J.T., Tizebatowek,J.R., Cruickshank,R.W., Gouzy,J., Brown,S.D., Elliot,R.M., Fleetwood,D.J., McCallum,N.G., Rossbach,U., Stuart,G.S., Weaver,J.E., Webby,R.J., de Bruijn,F.J., and Ronson,C.W.	
TITLE	Comparative sequence analysis of the symbiosis island of Mesorhizobium loti strain R7A	
JOURNAL	J. Bacteriol. 184 (11), 3086-3095 (2002)	
MEDLINE	21999272	
PubMed	12003951	
REFERENCE	2 (bases 1 to 150050)	
AUTHORS	Ronson,C.W.	
JOURNAL	Submitted (07-FEB--2002)	
COMMENT	Department of Microbiology, University of Otago, P.O. Box 56, Dunedin, New Zealand; MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824 USA; Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, Unité Mixte de Recherche (UMR) 215 Centre National de la Recherche Scientifique (CNRS), Institut National de la Recherche Agronomique, Chemin de Borde Rouge, BP27, F-31126 Castanet Tolosan Cedex, France, c.w.ronson@otago.ac.nz	
FEATURES	Location/Qualifiers	
source	1. 150050	
	/organism="Mesorhizobium loti"	
	/mol_type="genomic DNA"	
	/strain="R7A"	
gene	/db_xref="taxon:381"	
	198.1457	
CDS	/gene="ints"	
	198.1457	
	/function="elements of external origin; phage-related functions and prophages"	
	/note="Gene name confidence : probable"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="P4-TYPE SYMBIOSIS ISLAND INTEGRASE PROTEIN"	
	/protein_id="CAD31406.1"	
	/db_xref="GI:20803828"	
	/db_xref="GCA:068526"	
	/db_xref="SPTRMBL:O68526"	
	/translation="MEPPKALSDVKNCRNAPRISKLVKLSGGGLQLWVPTGSRLMRLARFGKQLALIGSPLISLAERARDAKRLILRGIPAKRCKQASADTRSIADVEDYVTKKEGRADRTITKVMIDPAVTPFNKSVREIDPATIILVARSVGRYSARLRSTIGSFRAVALTAARADVDTILRGALVGPVTTPRAATDGAAGLLRAINARFDGPTTPRAALKMALIPRRGRIARAEMFPFESAVMSIPEGKMKAPRRVPSRQAVSITSLRISDSGALVPPSVASRRPISDTINAAALRMKYGEEKATANGFRATASITLNEGCKMHPDAIEROLAHITESNDVRAVAPRHEWERIKOMQWADYLDSEKSDGVISIGHGKSTVTPSRDLR"	
gene	/complement (1845.3002)	
	/gene="msi002"	
CDS	/complement (1845.3002)	
	/gene="msi002"	
	/complement (1845.3002)	
	/gene="msi002"	
	/function="miscellaneous; unknown"	
	/note="Product confidence : hypothetical"	
	Gene name confidence : hypothetical	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="HYPOTHETICAL PROTEIN"	
	/protein_id="CAD31407.1"	
	/db_xref="GI:20803829"	
	/db_xref="SPTRMBL:O68527"	
	/translation="MRWDYSLTSDFTFMWRRLSAKKRVTLVIGRCPDVLTLETARLHLADISVMOLVDPNGDSDPKRELTATNHEGLKLVPAEKIVAAPIEIGPVGVAVPRNTKASIDRAGCDRDPDYVINDISAMPWALITAVSVLLFPLDILANRPSVNLHTTSVSADLGAAGKSLRDSVTVVGFSGHBERQTTQNMPRKFWFLYLEGQRDRHLHLIDHNPDEICPVIPTFRKPRRDEVGHRDITLFDKVEPANTILLSEYNPFAYRQILFAMERYKALKPMGGCKLFWSPLSKSLIGVILACYDKRFGKSSDRDVGIVPVEVAVVDPEONAEIADFEIHSILGIEWEERIPATASAATVAA"	

```
/transl_table=11
/product="HYPOTHEITICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD31411.1"
/db_xref="GI:20803833"
/db_xref="SPTREMBL:O8KG00"
/translation="MEFVTDVGEIRFVYRASEKPYGAFSNLFRREIEFEGERPATSE
HAYOGAKRKEPVKAMLAAPSPALLAAMAAGLYWVDVSPGMSRTRKPRMPCVLRAPK
TOHEDLRELIVSTSGARLIEATVNEVNRMLGMEVNGGGRNMLGELIMEIRAKLOND
VSKAA"
complement(7080..7988)
/gene="msi007"
/complement(7080..7988)
/gene="msi007"
/function="miscellaneous; hypothetical/partial homology"
/feature="Product confidence : hypothetical"
/feature="Gene name confidence : hypothetical"
/feature="Predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAD31412.1"
/db_xref="GI:20803834"
/db_xref="SPTREMBL:O8KG29"
/translation="MSADVVGTEFTVDRLYADSLAYSSEDLGSCGNVLISLAMI
RNVAPVSLGDEWVGRIVAVAFRAGATRIYRQTEHRSPTLQRLHPCGVHTFEF
CPETSIDLPRYEPIDMEVSRALPAIKNCNVFYDRLSGALVDAMEAADGTYF
PSALGDELFRRAVAAASVLFSAERIALTAQVAPHDALFLVTHGGALEVRGAL
KAMCKSYPAPELRDTCGSDWVSVGVIMLISRRDRARTRKISIDEFVVGIRAGQLAA
ENCAVVGARIGIFRERGAERYASVLAG"
complement(8041..9972)
/gene="msi008"
/complement(8041..9972)
/gene="msi008"
/function="miscellaneous; unknown"
/feature="Product confidence : hypothetical"
/feature="Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL TRANSMEMBRANE PROTEIN"
/protein_id="CAD31413.1"
/db_xref="GI:20803835"
/db_xref="SPTREMBL:O8KG28"
/translation="WTVPSKRLRPETIEILASAHAGDIAPALSQKRLVALIASG"
Query Match 6.1%; Score 142.2; DB 1; Length 150050;
Best Local Similarity 57.9%; Pred. No. 1.2e-19;
Matches 252; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
```

```
QY 653 AAATTAAGACTTTTCCGCAACGACATCTACTATCTTGATAGTAATAAACAATTG 712
DB 101139 GGACAGAGGTCTCTCCCTCCCAACGTTTCCACCAGACGGGTATCTCAGTGAACGATTT 101080
QY 713 CATGCGAGAGCTCA 727
DB 101079 TGTGACCCGAGCTGA 101065

RESULT 13
BX572601/c
LOCUS BX572601 348580 bp DNA linear BCT 18-DEC-2003
DEFINITION Rhodopseudomonas palustris CGA009 complete genome; segment 9/16.
ACCESSION BX572601 BX571963
VERSION BX572601.1 GI:39649375
KEYWORDS complete genomes.
SOURCE Rhodopseudomonas palustris CGA009
ORGANISM Rhodopseudomonas palustris CGA009
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Rhodopseudomonas.
1 (bases 1 to 348580)
Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L.,
Land, M.L., Pelletier, D.A., Beatty, T.J., Lang, A.S., Tabita, F.R.,
Gibson, J.L., Hanson, T.E., Torres y Torres, U., Perez, C.,
Harrison, F.H., Gibson, U., and Harwood, C.S.
Complete genome sequence of the metabolically versatile
photosynthetic bacterium Rhodopseudomonas palustris
Unpublished
2 (bases 1 to 348580)
Larimer, F.W. and Harwood, C.S.
Rhodopseudomonas genome consortium
Direct Submission
Submitted (24-JUL-2003) Submitted on behalf of the Rhodopseudomonas
Genome Consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA;
larimerf@ornl.gov

FEATURES
Location/Qualifiers
1..348580
/organism="Rhodopseudomonas palustris CGA009"
/mol_type="genomic DNA"
/strain="CGA009"
/db_xref="taxon:258594"
194..3916
/locus_tag="RPA2457"
194..3916
/locus_tag="RPA2457"
/function="InterPro IPR002048"
/feature="Observed by proteomics
Citation: Proteomics from Verberkmoes et al. (2003)
unpublished"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAE27898.1"
/db_xref="GI:39649376"
/translation="MPPQERILRDGAGHPREBPAPDARVOREAKMHPPTPOSDIAG
HDCNDPRHBEHNMHEHEEDARHARKILKRNKIGLRHIDWPSPRTTGAQWIKETLL
TIAVAGAVLIRGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV
GRIVAVLIRGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLVGLV
DPDQVTVASGSTRPLATGAPSGRPAPELPTQAPATPDAGAVAPGATAPSGSTPTP
PPADTSDTMGVALGEMDLSTLSIGLGDQNLNIGLNGNLVVDDOGRKRWTFEN
ISLIRPSGTGVALSVGEEGIKRWSLRVVGPAHDOVRSEILRANOVPTRTLLIAL
RLKULTYSADPEPLTGDUKGEYGRDGVPTTFRKGITAGAGTYIDSTPYPMAIDOTL
SFENDANRVLIAPEKVLAGEGRVTLTLSTLSPGKSVNDKMLGSGGTTIVFGAQNBS
PLIFNRILAVRVPDSANRVLITGADVNGSIVAGSSISIVSGSPRTLLIAGLQVMS
VSLIKRWMPALIVLELREWRVNDRIIAGSIQITDIIANSPMNLSRGPPIIDBGLVNV
ILGGAALRPVDGMPVVDADMRVTRGTATVAINQMANADTDGGRKVALSDILEFV
DLAPKPARIKFKLEGGVPAVAVLASGRSDAAALPIDNSSKGNVSAOVALGMPI
QRELTKDTSISITADLNGSPADKLAAGKLEANTLKYIANNQGYQVGVKYNQGA
```

```

misc_feature
/locus_tag="RPA2457"
/note="I1 probable transmembrane helix predicted
byTMHMM2.0"
complement (3886..3907)
/note="Transmembrane prediction, 98+%"
complement (4065..4625)
/locus_tag="RPA2458"
complement (4065..4625)
/locus_tag="RPA2458"
/function="InterPro IPR000516 COGs COG3038"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative cytochrome b561"
/protein_id="CAE27989.1"
/db_xref="GI:39649377"
/translacion="MTTTLDTADKGPQYGAVALIHWLIFPAVIALPASMVYAGL
EKTDPLRTLYDMHKAAGTAVLGLAVIRIMIKHGAPDLVPSPIREIVARISHGL
YLLALVPTGLGNTLCSRGVDLGLIPLAKNDATAGALHAHEPLFLTALVLIH
VIALHWQYPRHATIGRMVPMIRGK"
order/complement (4500..4565), complement (4398..4454),
complement (4269..4334), complement (4116..4181)
/locus_tag="RPA2458"
/note="4 probable transmembrane helices predicted
byTMHMM2.0"
complement (4846..6063)
/locus_tag="RPA2459"
complement (4846..6063)
/locus_tag="RPA2459"
/function="InterPro IPR007114 COGs COG0477"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative permease"
/protein_id="CAE27900.1"
/db_xref="GI:39649378"
/translacion="MDHRSRTGTFILGLCFVLILGRGHSPTVFLPISSFGMD
RAVVAIVSLTALCTGLASPLVGRILPRSGPRVYITGMILGALLILAAVARIWLT
QIVGVCVGLGIAFTGNVPNSILIGRRFKRILPANAAYVSATGATVLMPTAQILIT
DRYGMREAYVILIGAMVILIVPLSLMWRFPAGADHAQRPADADEGWTGSAR
HHAFFMALFATFFFTAIQMSISAQVAVIVADAGTPTLQATWAGFSGVIVYMLGVS
MLGVIGRRPSILFSVAISITGVLMLLMQVNPMLILGDFVICFGSGIRSGPLITA
TAMSIPIGRGVITGITSIGSLGSFSGWCGLLHDLTQSYNPVLGFALVAVLGM
IPLVAVPALRERO"
complement (5580..6063)
/locus_tag="RPA2459"
/note="Signal predicted by SignalP 2.0 HMM (Signal
peptide probability 0.601) with cleavage site probability
0.239 at residue 28"
order/complement (5980..6036), complement (5854..5919),
complement (5782..5832), complement (5686..5751),
complement (5584..5649), complement (5488..5553),
complement (5332..5397), complement (5221..5286),
complement (5149..5199), complement (5068..5133),
complement (4972..5028), complement (4861..4926)
/locus_tag="RPA2459"
/note="I12 probable transmembrane helices predicted
byTMHMM2.0"
complement (6148..7404)
/locus_tag="RPA2460"
complement (6148..7404)
/locus_tag="RPA2460"
/locus_tag="RPA2460"

```

```

/EC_number="6.1.1.1"
/function="InterPro IPR002305; IPR002307 COGs COG0162"
/note="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/unpublished=1
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="tyrosyl-CRNA synthetase"
/protein_id="CAE27901.1"
/db_xref="GI:39649379"
/translacion="MTAFKSDPMNVVQSRGFTIQIDPPSLDLAAKGEVAVAYGDC
TAASIVGHILSTMLHMLQATGNKPIALMGGFTVGDGSGDETRKILTYIQIDAN
KESIKGTFKFIKFGDSDALMADNAEMLTJKNLYEMLEIRHFSINMLTMDSVK
LRDRROELSFIEFNMILOSVDVEVLARRYKCNLQMGSDQGNITVGVLDGRMGT
HOLHALTCPLITTAAGKXMGKTAAGVAVLAPYDYQVXRENTDAVGRFLKLE
TILPMDEIRSLTQGAENINDAKRIILATEPAILMHGGEADKLETTARTTEGGCTQ
DPSVEISRGELDAGIGVAVFAEKGTVLANSGBARQIYAGGILKVNDAFVTDKMTL
TATDLSIDGVIKISMCKKHVLLRLA"
7562..8665
/locus_tag="RPA2461"
7562..8665
/locus_tag="RPA2461"
/function="InterPro IPR005338 COGs COG2377"
/note="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/unpublished=1
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Protein of unknown function UPF0075"
/protein_id="CAE27902.1"
/db_xref="GI:39649380"
/translacion="MMMTAIGLMSGTSLDGVVALIKTDGRVYALGPSGYRPTETE
RSILROALAAVOLTARDARPGVLAABEAVIAHAEAVAFVANOILSGSDIVFR
HGTVALRRPAKRTVTOIGDAKALAKAIRIPVYCDPRAADVAAQGNRPVYHRAIA
QSTGRDGPVGVNIGVSNVTVYDGTITLACDTGKNALDDPFRITLGRFDCER
LASQGVLDQAKTEALQHPFAKPTKSLDRNDFANLTLDWSPADGAATLTAFA
IAVQVLPLEKPPQSWIVTGGANRILTMMLREALA PASVESDALGWSADAMEAQAF
GLATRGKGLPLSYPATIGVAPMTGILLAR"
complement (8720..9367)
/locus_tag="RPA2462"
complement (8720..9367)
/locus_tag="RPA2462"
/function="InterPro IPR000379 COGs COG3945"
/note="observed by proteomics
Citation: Proteomics from VerBerkmoes et al. (2003)
unpublished"
/unpublished=1
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="conserved unknown protein"
/protein_id="CAE27903.1"
/db_xref="GI:39649381"
/translacion="MPBVITFTGPAKLEGRHYHAKQKNAPVAVLHHPQGRMNP
IVQVYVAFARQPSVLRNFRGVGSQSPDGTGBLADASALDAQITNIPARC
WVAGFSGMIGQMILMRREPVGSIAPENRYPFSLPAPSSGLIVGKEDIVA
PAKVNVIVKMLTQKGVIVDQOTIPGANHFEDRMEPLMETITSYDMLANVR"
9605..10777
/locus_tag="RPA2463"
9605..10777
/locus_tag="RPA2463"
/function="InterPro IPR000192 COGs COG1104"
/codon_start=1

```

Query Match 6.1%; Score 142.2; DB 1; Length 348580;
Best Local Similarity 57.0%; Pred. No. 1.1e-19;
Matches 280; Conservative 0; Mismatches 208; Indels 3; Gaps 1;

CY 244 TGAACCTTACCTTCATCTGTCATATCATGCGCTTTTAAAGGTTGACCTTGAAT 303
|||||

Db 122447 TATAGTTTTCCGCGGCGGATCAGCAATTAGGCCCGGAGACGATGGCGCGCTCCGGA 122388

Qy 304 CCTATCTGACCATTAATATTTTTAGATGTTTCAATAGAAAGCTC--TGAATCAACG 360

Db 122387 CCGAGTTGGCCCGCAAAATGTTCCAAACCTGCTGATTTGCAAGCAGCGGTAGTTAG 122328

Qy 361 GACTGGGAAATTAAGTGTGATATCCAGAAATTTGTCATCTTCAAGTAAACACCTCAGCA 420

Db 122327 GCGTCAATTCACAGATGATAGAAACCGACCTTGTGACCTCAAGCCGACACTTCCCGG 122268

Qy 421 GTTAAACACCTAAGTTCTCACCAGATGTCTCAATATCCGACGGAATATTTATTTGCT 480

Db 122267 GTCAAGATGCCACATCTTCAACCAAAAGTCTCGACGTTGGCGCTCCAGCCGCGAC 122208

Qy 481 TCTCTTGAACCGTAGACCTTTCCACATGCAAGATTTGGAACCTCTTGAAGTACTACTG 540

Db 122207 GCGCGCGACGCGGAGTTTCCAGGACAGAGCTTAGGCACTTCTGAAAGCACTACCG 122148

Qy 541 GAATGATTTGCAATATTTGCTACACCTTGGCTGATGAGTAAAGTCCGTTAATGTTGCT 600

Db 122147 GAATCGCTGCAACAGATGCGCACCGCTTGCCTTGAAGAAAGATCGGACAGCGCCGA 122088

Qy 601 AAAAAGCAGAGAGCAAGATGATGAGATGACCTCTGTTCAATGCAATAACTAAT 660

Db 122087 ATGAAGCCGAGAGAGAGCGGAGGATGAAATGCGCTCGGCGCTCTGCAAGTACAG 122028

Qy 661 GACTTTTGGCCACAGCATCTAATCTTGTGATAGTAAATAAACAATTCATGTCGA 720

Db 122027 GTCGCTGTGTCAGCAGCTTCAACGAGGCGGTCATGTTAGAAAGCAATGCGGCGCC 121968

Qy 721 GAGCTCATTCG 731

Db 121967 GAATCAATCG 121957

RESULT 14
SLHAUR/c 2365 bp DNA linear BCT 09-JUN-1995

LOCUS SLHAUR 2365 bp DNA linear BCT 09-JUN-1995

DEFINITION S.lividans haur gene.

ACCESSION X77574

VERSION X77574.1 GI:455520

KEYWORDS ORF; resistance protein.

SOURCE Streptomyces lividans

ORGANISM Streptomyces lividans

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE 1 Zhou, X., Deng, Z., Hopwood, D.A. and Kieser, T. Streptomyces lividans 66 contains a gene for phage resistance which is similar to the phage lambda eads endonuclease gene Mol. Microbiol. 12 (5), 789-797 (1994)

JOURNAL MEDLINE 94328930

PUBMED 8052130

REFERENCE 2 (bases 1 to 2365)

AUTHORS Kieser, T.

TITLE Direct Submission

JOURNAL Submitted (07-FEB-1994) T. Kieser, John Innes Institute, Colney Lane, Norwich NR4 7UH, UK

FEATURES

source 1..2365

location/Qualifiers

1..2365

/organism="Streptomyces lividans"

/mol_type="genomic DNA"

/strain="66"

/db_xref="taxon:1916"

1039..2358

/gene="haur"

1039..2358

/gene="haur"

/codon_start=1

/transl_table=11

/product="phage Haur3 resistance protein"

/protein_id="CAA54681.1"

/db_xref="GI:455521"

/db_xref="GOA:O54346"

/db_xref="SPTRMBL:O54346"

/translation="MRCHNGRSVAVGAKVIGFPMGEGRTGLPERFALPDPSLGLADEDYATLRDPEDEETRLTILSGIDVAAYPELFEQALSEVMDSILRGDTDTVRTOYRIVAGPFLPSFHRKYQOAGSGAPTELKVDPEKNPNSIHAVGNSGVKTVILHKIHAHTLNSRNRSSSWLEDRDLHPETNIVVYSFSAFDSHTPQNVAVNRVGDVDYQVYGLKTDDEEVKYNALGAFAPRVERCEVEDHPKARMGDVLAKLEETDPLFHLIRELAQDRDPDRPRVEFGSSGKIVLTLARIVQHTERTLVILDEPEAHLPPLTSPVRLTSELRLDRNGLAVIATHSPVVAOETPREAVMLRRAGDRLVDHPELETPEGENVITREIFGLEVRTGFNRLIQLLADEGYVFRNP"

ORIGIN

Query Match 4.5%; Score 104.8; DB 1; Length 2365;

Best Local Similarity 53.7%; Pred. No. 1.2e-11;

Matches 217; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

Qy 339 GAATGGTTTCAATAGAAAGCTCTGATATCAAGGACTGGATTAAGTGTGTATCCG 388

Db 2353 GATTTCTTGAAAGACATACCTTGTGCGGAGAGAACTGATCAACCGTTAAACCTCG 2294

Qy 389 AATTTGTCATTCAGTAAACACCTCAAGTTAAACACCTAAGTTCTCACCGAATG 448

Db 2253 TCGCGGACCTCCAAATCCAAAGATCTCCGGGTATCACTCCAGCTTCTGCCGAATG 2234

Qy 449 TCTCAATATCCGAGCAGATATATTATTATCTTCTTCTTGAACCTTGAAGCTTCCACATG 508

Db 2233 TCTCGATCTCGGGGATGATCCACGACAGTCAATCCCGCTCTGCGGAGGCGCCAGACGG 2174

Qy 509 AGGATTTGGAACCTCTGCACTACTCGGGGATGAGTGCATTAATTTGTACACCAT 568

Db 2173 CCTACGAGGCGTCTCTGAGAACGACTGGGAGTGGTACGATTAACCGAAGCGCT 2114

Qy 569 TCGGTGATCGAGTAAAGTGTGTTATGTTCTGTAATAAAGCAGAGCAAGGTGATGCA 628

Db 2113 TTGGTCTCTAGACAGCTCAAGCAAGCGTCCGACAGAGTGCAGCAAGCGGCGGTGA 2054

Qy 629 GATGAACCTCTGCTTCATCGAATAACTAATGATTTTCCGCAACGACATCTACTAATC 688

Db 2053 GATGGCGCTCCGGTTCATCGATCAGAGAGAGGTGCGTGTGTGTGTACCAAGCC 1994

Qy 689 TTGTATGTAATTAATTAATTAATGATGCTCCAGAGCTCATTTGA 732

Db 1993 GGGCCAGATTNAGCAGACGATCTTGTCTCCGAACTCAGCCCA 1950

RESULT 15
AR204117 10523 bp DNA linear PAT 20-JUN-2002

LOCUS AR204117 10523 bp DNA linear PAT 20-JUN-2002

DEFINITION Sequence 13 from patent US 6365723.

ACCESSION AR204117

VERSION AR204117.1 GI:21500681

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 10523)

AUTHORS Blattner, F.R., Burland, V., Perna, N.T., Plunkett, G. and Welch, R.

TITLE Sequences of E. coli O157

JOURNAL Patent: US 6365723-A 13 02-Apr-2002;

FEATURES

source 1..10523

location/Qualifiers

1..10523

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

Query Match 3.2%; Score 75; DB 6; Length 10523;

Best Local Similarity 80.4%; Pred. No. 1.8e-05;

Matches 115; Conservative 0; Mismatches 20; Indels 8; Gaps 2;

Qy 2168 CCGCTCTCGCTCATACAGACATTCACCTACAGTTATGCGGAGAGATATGATGCTGCG 2227

Db 10258 CCAATCTTGTGCTCAATACAGGATTCACCTTACGTATGCGAGAAAGATATGATGCTGCA 10317

QY	2228	TTGCGGAGTCGTGAAAGAAAGAGTCACTGCGTGTGACATCACTGCTATCTTC	2287
Db	10318	-----GAAGTCGTGAAAGAAAGAGTCACTGCGTGTGACATCACTGCTATCTTC	10371
QY	2288	TTACTGTGTATGCAAGTCGTAGT	2310
Db	10372	--ATTGCTATGCAAGTCGTAAAT	10392

Search completed: October 6, 2004, 07:57:31
Job time : 9055 secs